

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CURSO DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO

Marcus Vinicius Pavan

SestatNet: Módulo de Análise por Componentes
Principais.

Dissertação submetida à Universidade Federal de Santa Catarina como parte dos requisitos para a obtenção do grau de Mestre em Ciência da Computação

Prof^ª. Silvia Modesto Nassar.

Florianópolis, Março de 2003.

Módulo de Análise por Componentes Principais.

Marcus Vinicius Pavan

Esta Dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação Área de Concentração Sistemas de Conhecimento e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

Prof. Fernando Álvaro Ostuni Gauthier

Banca Examinadora:

Prof^a. Silvia Modesto Nassar
Presidente da Banca (orientador)

Prof^a Araci Hack Catapan

Prof. Pedro Alberto Barbetta

Prof. Marcelo Menezes Reis

SUMÁRIO

1. Introdução.....	10
1.1. Objetivos.....	14
1.1.1. - Objetivo geral.....	14
1.1.2. - Objetivos específicos.....	14
1.2 – Sistema de Ensino-Aprendizagem de Análise Estatística de Dados - Sestat	15
1.2.1 – SEstat v.1.0 – A criação do projeto.	15
1.2.2. - Funcionamento do Sestat	16
1.2.3. Arquitetura do SEstat	16
1.2.4. Melhoramentos na versão 1.0 do SEstat.....	19
1.2.5. – SEstat v.2.0.....	19
1.2.6. – SESTATNET.....	20
2.1 Análise Estatística Multivariada.....	21
2.2 – Embasamento teórico da ACP.....	22
2.2.1. REDUÇÃO DA DIMENSIONALIDADE DOS DADOS	23
O problema em \mathbb{R}^p	24
O problema para \mathbb{R}^n	27
O Modelo	30
Os Problemas de Escala.....	34
Transformação de Resultados.....	36
A Matriz de Escores Fatoriais	37
3 – A Inteligência Artificial e a Educação.....	40
3.1 - O Construtivismo e a Educação Tradicional.....	40
3.2 – A Educação a distância.....	43
3.3 - Inteligência Artificial.	47
3.4 - Sistemas Especialistas	53
3.4.1. - A Base de Conhecimento	55
3.4.2 - A Máquina de Inferência.....	56
3.4.3. - A Interface com o Usuário	57
3.5 - Sistemas Especialistas Apoiando a Educação.....	57
4 – DESENVOLVIMENTO DO MÓDULO ACP	61
4.1 - Java Server Pages.....	61
4.2 - Classes do SESTATNET	63
4.2.1 – A Classe Administrador.....	64
4.2.2 – A Classe Estatística	67
4.2.3 – A Classe SETip.....	69
4.2.4 - Interface do SESTATNET	69
4.3 – O Desenvolvimento do módulo ACP	72
4.3.1 - Validação dos algoritmos.....	73
4.3.2. INTERFACE DO ACP.....	74
5. – Conclusão.	86

6 – Referências Bibliográficas.....	88
ANEXOS	93
Algoritmos	93
Algoritmo em Portugol para a determinação de AutoValores e AutoVetores	93
Algoritmo em Portugol para determinação dos Escores fatoriais	95
Algoritmo em Portugol para determinação da matriz de Correlação	96
Métodos comuns ao SESTATNET	97
Métodos exclusivos do Módulo de ACP	100

LISTA DE FIGURAS

Figura 1 - Arquitetura do SEstat v.1.0.....	19
Figura 2 - Uma visão conceitual dos sistemas de Inteligência Artificial. (SCHUTZER 87)	49
Figura 3 - Estrutura convencional de um Sistema Especialista (CHAIBEN 03)	55
Figura 4 - Requisição de uma página JSP.	62
Figura 5 - Núcleo do Sistema SESTATNET.....	63
Figura 6 - Classe Administrador.....	64
Figura 7 - Classe Variável	66
Figura 8 - Classe Educador.....	66
Figura 9 - Classe Estatística	67
Figura 10 - Classe Relatório	68
Figura 11 - Classe SEtip	69
Figura 12 - Entrada do sistema. (NAKAZAWA e MARAFON 2003).....	70
Figura 13 - Escolha da base de dados do aluno. (AKAZAWA e MARAFON 2003).....	71
Figura 14 - Tela de Apresentação integrada ao Módulo de ACP	74
Figura 15 - Seleção de Variáveis.....	75
Figura 16 - Tipificação de Variáveis	76
Figura 17 - Seleção da Variável Suplementar	77
Figura 18 - Aviso do Sistema em caso de escolha errada	77
Figura 19 - Tela de apresentação de características descritivas	78
Figura 20 - Matriz de Correlação	79
Figura 21 - Autovalores - Valores numéricos e gráfico de fatores.....	80
Figura 22 - Cargas Fatoriais	81
Figura 23 - Coeficientes de Escores Fatoriais	81
Figura 24 - Cargas Fatoriais - Fator 1 x Fator 2	82
Figura 25 - Cargas Fatoriais - Fator 1 x Fator 3	82
Figura 26 - Cargas Fatoriais - Fator 1 x Fator 2 x Fator 3.....	83
Figura 27 - Escores Fatoriais dos Indivíduos	84

LISTA DE QUADROS

Quadro 1 - Comparativo entre as salas de aula Tradicional e Construtivista, segundo (Brooks e Brooks, 1995).....	42
Quadro 2 - Comparação entre a Inteligência Artificial e a Inteligência Natural. (CHORAFAS-88).....	51

ABSTRACT

The artificial intelligence is an area intimately related to the process of human learning. One of its several areas, the one of specialist Systems, gives capabilities of a human specialist to a computing system increasing exponentially its capabilities. A human specialist has its capacities usually conditioned to his own human condition, i.e., analysis capability influenced by external factors and a limited capacity of storage and recovery of great masses of information. The specialist system is exactly a tool that offers to the human specialist those two resources, in other words, capability of complex analysis and storage of great volumes of data, given that the later has its capabilities flagrantly limited if compared to the capacities of the computer. The creation of such specialist systems towards the education area it has been showing great research possibilities and practical results. The use of technology resources associated to the practices of teaching-learning of the E-Learning it is already a reality. Invariably those resources have in its fundamentals, techniques of a specialist system.

The aggregation of the several specialists' knowledge in Statistics, together with an open interface for Internet, it is basic subsidy for the construction of the module proposed in this research. This module will integrate into SestatNet (System of Specialist of Teaching-learning of Statistics) that is a system developed at Applied Statistics Laboratory - LEA - in the Federal University of Santa Catarina - UFSC.

This module approaches the technique multi-varied called analysis by main components providing to the student the knowledge and statistical procedures to evaluate the association among a group of quantitative variables, following the pedagogic bases that orientate the project of the system SESTATNET.

Key words: Specialist systems, Artificial Intelligence, Statistical Analysis Multi-varied, Analysis by main components.

"O temor do Senhor é o princípio do conhecimento; mas os insensatos desprezam a
sabedoria e a instrução."
Provérbios 1:7, A Bíblia Sagrada

Resumo

A inteligência artificial é uma área intimamente relacionada ao processo de aprendizado humano. Uma das suas diversas áreas, a de Sistemas especialistas, confere capacidades de um especialista humano a um sistema computacional aumentando exponencialmente suas capacidades. Um especialista humano tem as suas capacidades normalmente condicionadas a sua própria condição humana, ou seja, capacidade de análise influenciada por fatores externos e uma capacidade limitada de armazenamento e recuperação de grandes massas de informação. O sistema especialista é justamente uma ferramenta que oferece ao especialista humano esses dois recursos, ou seja, capacidade de análises complexas e armazenamento de grandes volumes de dados, visto que este tem suas capacidades notoriamente limitadas se comparadas às capacidades do computador. A criação de tais sistemas especialistas voltados para a área de educação tem mostrado grandes possibilidades de pesquisa e resultados práticos. A utilização dos recursos da tecnologia associados às práticas de ensino-aprendizagem do Ensino a Distância já é uma realidade. Invariavelmente esses recursos têm em seus fundamentos técnicas de um sistema especialista.

A agregação do conhecimento de vários especialistas em Estatística, juntamente com uma interface aberta para a Internet, é subsídio básico para a construção do módulo proposto nessa pesquisa. Esse módulo irá integrar-se ao SestatNet (Sistema de Especialista de Ensino-Aprendizagem de Estatística) que é um sistema desenvolvido no Laboratório de Estatística Aplicada – LEA- na Universidade Federal de Santa Catarina - UFSC.

Este módulo aborda a técnica multivariada chamada de análise por componentes principais provendo ao aluno os conhecimentos e procedimentos estatísticos para avaliar a associação entre um conjunto de variáveis quantitativas, seguindo as bases pedagógicas que norteiam o projeto do sistema SESTATNET.

Palavras chaves: Sistemas Especialistas, Inteligência Artificial, Análise Estatística Multivariada, Análise por componentes principais.

1. Introdução.

Como o ensino-aprendizagem continua em constante modificação e evolução, o advento do computador e principalmente da Internet, vem facilitando o acesso e a possibilidade de disponibilizar conhecimento e informação com muito mais velocidade e em tempo menor estão quebrando paradigmas e permitindo que se conheça e que se aproveitem novas formas de acesso ao conhecimento.

O SESTAT – Sistema Especialista para Ensino-Aprendizagem de Estatística, desenvolvido no LEA (Laboratório de Ensino de Estatística Aplicada – INE – UFSC) direcionado a integrar um ambiente de Ensino-Aprendizagem em Análise Estatística de dados, foi concebido pela ausência de uma ferramenta capacitada tanto como uma base de conhecimento confiável e também como um ambiente em que o aluno pudesse ser um agente integrado ao processo de sua aprendizagem. Inicialmente o SESTAT continha módulos para análise univariada, análise bivariada e inferência estatística. Esta pesquisa é uma das primeiras a contribuir com um novo módulo para o SESTAT, na área de análise multivariada, mais especificamente a técnica de Análise por Componentes Principais (ACP).

Ao mesmo tempo em que esta pesquisa tem o seu direcionamento voltado para o desenvolvimento de um dos módulos de análise multivariada do SESTAT, um outro módulo também está sendo desenvolvido pela aluna Jaqueline Piveta, sendo esse voltado para a Análise de Correspondência Múltipla. (PIVETA, 2003)

O SESTAT foi inicialmente concebido para ser um sistema que auxiliaria aulas presenciais tendo como recomendações a implantação de módulos que abordassem os métodos de análise multivariada e migrasse para a modalidade de educação a distancia, passando a ser chamado de SESTATNET.

A **APC** é uma técnica hoje muito utilizada. A aprendizagem dessas técnicas regularmente exige o conhecimento específico de um professor e um programa (software) estatístico comercial. A contribuição esperada deste módulo de análise por componentes principais, para o SESTATNET visa aprimorar as formas dessa aprendizagem, sendo presencial ou a distância, ao utilizar-se uma ferramenta que pode oferecer conceitos

pertinentes a Análise por Componentes Principais, realizar cálculos resultando em valores numéricos e gráficos, e ainda integrados a um ambiente com inteligência artificial será oferecida ao aluno uma forma muito mais dinâmica, rica e produtiva de aprendizagem. O SESTATNET não tem os limitadores físicos de outros softwares estatísticos, pois é totalmente utilizado via Internet. A concepção do módulo **APC** tem a mesma preocupação daqueles que iniciaram o projeto há alguns anos, ou seja, todas as classes utilizadas, produzidas ou utilizadas são de domínio público. Com isso o SESTATNET com o **APC** poderá representar uma nova forma e uma nova opção no ensino-aprendizagem em técnicas de análise estatística multivariada.

Relevância da Pesquisa

A pesquisa de novos algoritmos que pudessem propiciar os resultados necessários para a **APC** foi feita utilizando-se da literatura disponível e principalmente a Internet.

Foi em (JYNCH 2003) que se pôde encontrar vários algoritmos produzidos pelo pesquisador John M. Lynch do Barrett Honors College, Arizona State University. LYNCH, que produziu uma série de algoritmos utilizando uma linguagem não estruturada (GW-BASIC). A partir desses algoritmos foram realizadas sucessivas traduções dessa linguagem para linguagens mais estruturadas até poder ter protótipos estruturados dos métodos em JAVA.

Nessa fase foi possível submeter o algoritmo a vários testes onde ele se mostrou eficiente.

A validade do algoritmo foi confirmada por testes feitos comparativamente entre o algoritmo produzido e o software Statistica v.5, que analisando amostras de dados iguais chegaram aos mesmos resultados.

A utilização de algoritmos computacionais é fase necessária para a criação de sistemas seja para qualquer finalidade. Implementações relativas a análises estatísticas normalmente exigem conhecimentos matemáticos e principalmente de álgebra linear. Junto a técnicas que não são tão conhecidas publicamente às técnicas de análise multivariada tornam-se aplicações complexas e de difícil desenvolvimento. A utilização de pacotes

estatísticos (neste caso, o Statistica) para a validação dos algoritmos se faz necessário justamente por essa complexidade. Ao desenvolver-se um algoritmo e submetê-lo a uma linguagem de programação a chance de ocorrerem erros, que normalmente são particulares pela linguagem escolhida, acabam por prejudicar o andamento de testes e aproximações. Assim a validação por softwares tidos como eficazes tem sua importância nessa pesquisa. Ao apresentar resultados dessas validações também se pode verificar a diferença existente entre alguns softwares estatísticos que também abordam a **APC**. As leituras que se fizeram necessárias ao longo dessa pesquisa também revelaram que os autores da área da **ACP** alertam sobre esses diferentes resultados obtidos em softwares estatísticos. Não se obteve objetivamente uma causa ou explicação que justificassem essas diferenças, podendo ser esse uma pesquisa futura de caráter de avaliação sobre esses softwares.

Estrutura da dissertação

No capítulo 1 a apresentação da pesquisa. O principal interesse é mostrar quais são os objetivos, como será o funcionamento e também qual é o seu escopo de atuação.

No capítulo 1 também é apresentado o SEstat com um breve histórico do projeto, uma simples descrição de funcionamento do projeto, suas capacidades, propriedades e melhorias do sistema. A apresentação do SESTATNET também é abordada.

Como início da conceituação teórica o capítulo 2 aborda a análise estatística multivariada e também da técnica que é tema desta pesquisa a análise por componentes principais.

No capítulo 3 um aprofundamento nos fundamentos sobre Educação, Inteligência Artificial e Sistemas Especialistas. O tema é discorrido sobre a educação, seus métodos tradicionais e modernos, a teoria construtivista e ainda os aspectos teóricos da inteligência

artificial, mais especificamente os aspectos de sistemas especialistas e a aquisição de conhecimento.

No capítulo 4 o desenvolvimento do módulo proposto foi descrito formalmente, bem como o funcionamento dele em relação aos demais módulos do SESTATNET

Em anexo procura-se descrever os algoritmos que utilizados para a realização dos cálculos utilizados na análise por componentes principais e na matriz de correlação. As classes desenvolvidas que foram utilizadas para a criação dos métodos implementados e as interfaces.

1.1. Objetivos

1.1.1. - Objetivo geral

Desenvolver um módulo inteligente de Análise por Componentes Principais (ACP) seguindo a abordagem pedagógica do Sestat.Net – Sistema Especialista para ensino-APRENDIZAGEM de estatística utilizando a internet.

1.1.2. - Objetivos específicos

- Investigar os algoritmos matemáticos e computacionais da **ACP**.
- Investigar a modularização de sistemas sob a ótica de Sistemas Especialistas
- Pesquisar tecnologias WEB para a melhor integração da ACP ao **SESTATNET**
- Desenvolver um módulo que integre os conhecimentos adquiridos e torná-lo integrante do SESTATNET.
- Validar os algoritmos produzidos do ACP.

1.2 – Sistema de Ensino-Aprendizagem de Análise Estatística de Dados - Sestat

1.2.1 – SEstat v.1.0 – A criação do projeto.

O desenvolvimento deste item terá por base (DIAS 2000).

A criação efetiva do SESTATNET teve início com o trabalho de conclusão de curso em Ciências da Computação de Kirliam Maciel Dias ,em dezembro de 2000.

Como os objetivos da primeira versão do SESTATNET era, e ainda é, o de oferecer apoio ao ensino-aprendizagem de análise estatística de dados. Para que esse objetivo fosse alcançado, foram determinadas algumas características que seriam necessárias ao sistema.

Base de dados flexível : Permitir ao usuário trabalhar com qualquer base de dados desejada. Dessa maneira, o SEstat dá ao usuário a possibilidade de usar vários conjuntos de dados da sua área de conhecimento. Esta característica proporciona ao usuário a *generalização* do conhecimento estatístico.

Ser uma ferramenta de análise estatística de dados: Além de recomendar um método estatístico adequado para determinada análise, o SEstat também aplica aquele método e mostra ao usuário os resultados estatísticos obtidos. O processo de seleção do método estatístico inclui a verificação das suposições necessárias para a sua aplicação, tais como normalidade, homocedasticidade, nível de mensuração das variáveis, por exemplo. Isso permitirá que o usuário atinja os níveis de conhecimento de definição e conceituação.

Disponibilizar o caminho que está sendo percorrido: Mostrar ao usuário o caminho que uma dada interação percorre até chegar ao resultado estatístico obtido, e também os caminhos que o sistema pode seguir no caso de respostas diferentes. Essa característica tem como objetivo localizar o usuário dentro do raciocínio estatístico envolvido, e permitir que ele/ela desenvolva a capacidade de generalização do conhecimento estatístico.

Help sensível ao contexto: Possibilitar ao usuário, a qualquer momento da interação, a opção de acessar informações a respeito das questões que o sistema lhe propõe. Proporciona ao usuário reconhecer e apreender conteúdo estatístico.

1.2.2. - Funcionamento do Sestat

Na sua versão inicial, o sistema foi desenvolvido através da linguagem de programação Object Pascal no ambiente DELPHI 3.0™, e teve como característica marcante a comunicação com o *software* Statistica 5.0™, cujo papel é o de realizar o processamento da análise estatística de dados propriamente dita. (DIAS 2000)

A principal característica que diferenciaria essa versão das demais é a dependência do *software* Statistica para toda a produção de resultados numéricos e gráficos. O SEstat funcionava inicialmente como uma SHELL inteligente dirigindo o aluno durante as fases de análise e enviando e recuperando conhecimento do *software* Statistica.

Partindo de uma base de Dados (em formato .dbf) fornecida pelo usuário, o SEstat faz um conjunto de perguntas a ele, e fornece ao mesmo suporte para que essas sejam respondidas corretamente.

Ao longo da interação do usuário com o sistema, o caminho a ser percorrido nesta interação é escolhido através das respostas dadas pelo usuário às perguntas referentes à análise estatística que se está realizando. No final da interação, o sistema escolhe e aplica uma técnica de análise estatística para os dados fornecidos pelo usuário, e explica ao mesmo por que aquela técnica foi escolhida, bem como apresenta explicações que possibilitam a interpretação dos resultados estatísticos encontrados. (DIAS 2000)

1.2.3. Arquitetura do SEstat

O conhecimento estatístico foi dividido e distribuído entre módulos (objetos), dando assim a cada módulo uma característica e uma tarefa específica. Os módulos

possuem a capacidade de se comunicarem entre si e de decidir qual módulo é mais capaz de executar determinada tarefa.

São eles: o Gerenciador, a Estratégia, o Processamento, o *software* Statística 5.0TM, o Mecanismo de Explicação e a Interface.

A seguir estão descritos estes módulos do sistema e suas respectivas funções:

Gerenciador: Decide qual módulo é capacitado para executar a tarefa que é necessária em um determinado ponto da execução da análise de dados.

Estratégia: De acordo com as informações fornecidas pelo Gerenciador, este módulo irá procurar uma estratégia adequada para a resolução do problema em questão.

Processamento: Responsável pela comunicação com o *software* Statística 5.0TM, este módulo possui conhecimento sobre os passos necessários para a execução de um determinado procedimento estatístico.

Software Statística 5.0TM: O Statística realiza a análise de dados propriamente dita através da execução dos passos fornecidos pelo Processamento.

Mecanismo de Explicação: Consiste basicamente no módulo de ajuda do sistema, que tem por finalidade explicar ao usuário conceitos estatísticos, e é capaz de mostrar o caminho que foi percorrido para se chegar às decisões tomadas. Os conceitos estatísticos disponíveis no SEstat são de dois tipos:

a) Básicos – Tais como nível de significância, escala de mensuração de variáveis, variáveis dependentes, variáveis independentes, etc.

b) Interpretativos – Que apóiam o usuário a interpretar os resultados estatísticos oferecidos pelo SEstat.

Interface: Recolhe do usuário as informações que serão processadas pelo sistema, como por exemplo o nome e a mensuração das variáveis a serem analisadas. Mostra ao usuário a evolução do sistema, assim como as soluções obtidas e as respostas às perguntas feitas pelo usuário.

Módulo de Treinamento: Devido ao fato de trabalhar com uma base de dados flexível, o sistema não possui o controle e o conhecimento das características das variáveis com as quais o usuário está lidando; esta flexibilidade ao domínio pesquisado (“free context”) tem a desvantagem de permitir que o usuário responda a perguntas de forma errada sem que haja a intervenção do sistema para a correção da resposta. Em consequência disso, o sistema pode tomar um caminho errado na busca dos resultados estatísticos, fazendo com que o aluno não aprenda a realizar adequadamente a análise de dados.

A solução encontrada para resolver este problema foi a construção de um módulo de treinamento dentro do SEstat. Este módulo trabalha com uma base de dados fixa (já conhecida pelo sistema), e dá ao sistema a possibilidade de identificar erros nas respostas do usuário assim como avisá-lo sobre esses erros. A cada interação do usuário com o SEstat, o módulo de treinamento irá intervir fazendo observações a respeito da resposta que foi dada, comentando os acertos e fornecendo explicações para os erros.

Desta forma, o usuário que não estiver habituado a utilizar o sistema, utilizará o Módulo de Treinamento para aprender os conceitos estatísticos envolvidos e depois poderá aprimorar essa aprendizagem utilizando uma base de dados qualquer escolhida por ele.

O módulo de ajuda contém um dispositivo de acompanhamento sobre as respostas de cada usuário, ou seja, os erros e os acertos do usuário são contabilizadas em arquivo de dados. Assim, o professor pode comparar qual foi o desempenho dos alunos dentro de uma mesma turma, como também pode comparar o desempenho entre turmas distintas.

Esse acompanhamento possibilita localizar quais os assuntos da disciplina que mais oferecem resistência aos alunos e dá ao professor a oportunidade de replanejar a abordagem pedagógica utilizada para o ensino desses assuntos. Além disso, também permite que o professor personalize o ensino de Estatística para diferentes cursos e suas respectivas áreas de conhecimento.

Na figura 1 pode-se mostrar a relação entre os módulos que fazem parte da estrutura do SEstat v.1.0.

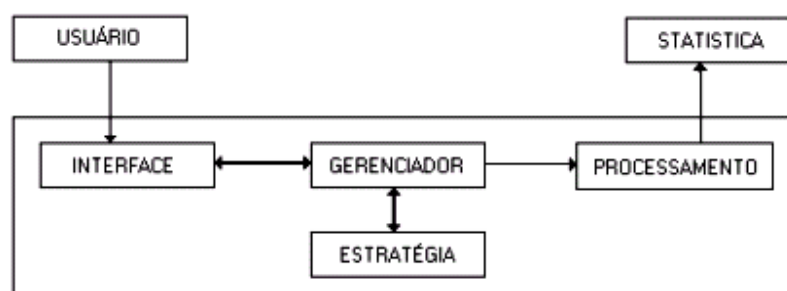


Figura 1 - Arquitetura do SEstat v.1.0

1.2.4. Melhoramentos na versão 1.0 do SEstat.

- a) A criação de módulos independentes e inteligentes que pudessem implementar todos os algoritmos necessários para as análises e demonstração de resultados estatísticos, portanto sendo independentes do software Statistica;
- b) O incremento na base de conhecimento;

1.2.5. – SEstat v.2.0.

A nova versão (inicialmente chamada EDUST) do SEstat apresentou melhorias em sua interface e também no aspecto de desempenho.

Essa versão já apresentou autonomia sobre os algoritmos das análises e também quanto a apresentação gráfica dos resultados, realizando melhoras significativas.

O aumento na base de conhecimento aconteceu com a criação do AMBIENTE COLABORATIVO PARA ENSINO DE ESTATÍSTICA COM O SESTAT, pelo estudante da pós-graduação RENATO BICA NOAL.(NOAL 2002).

1.2.6. – SESTATNET

A versão anterior do SEstat já havia produzido ótimos resultados, tanto do ponto de vista pedagógico como computacional. Mas ainda faltavam melhorias quando a forma de acesso e expansibilidade da sua base de conhecimentos.

A proposta apresentada por (NAKAZAWA e MARAFON 2003) trouxe o SEstat para a plataforma web, ou seja, o SEstat agora não possuía mais restrições quanto a plataforma de funcionamento, tampouco restrições quanto ao local de utilização. Sendo um sistema totalmente implementado em ambiente web agora o SEstat incorpora não só os conceitos de ferramenta de auxílio a aulas presenciais, mas também conceitos de ensino à distância.

Sendo uma interface totalmente reformulada, mas mantendo os mesmos princípios pedagógicos do projeto inicial, o SESTATNET apresentou um aumento na sua base de conhecimento. Devido a sua estrutura, definida durante a sua construção primária, o acréscimo de novos módulos na base de conhecimento ficou viabilizada.

Os módulos de análise multivariada foram criados focando dois métodos específicos: a análise de correspondência múltipla (ACM) e a análise por componentes principais (ACP).

A ACP é o objeto dessa pesquisa, por isso no próximo capítulo apresenta-se uma fundamentação teórica sobre esse método de análise estatística.

2.1 Análise Estatística Multivariada

A revisão da literatura e conseqüente fundamentação teórica são fundamentais e necessárias ao desenvolvimento deste trabalho e inicia-se por uma breve abordagem dos métodos estatísticos multivariados, dando ênfase à Análise de Componentes Principais.

A análise multivariada tem sido ultimamente muito utilizada. Em meios acadêmicos sempre foi uma interessante ferramenta para pesquisadores que necessitavam, a princípio, de procedimentos para análise multivariada de dados.

Na atualidade, as técnicas de análise multivariada ocupam cada vez mais espaço em atividades empresariais que envolvam o comportamento do consumidor, fundamentalmente por conferirem maior precisão aos resultados e agilidade na mensuração da satisfação de consumidores. (NIEDERAUER 2003)

Os vários softwares estatísticos disponíveis também tornaram tais técnicas acessíveis, mas ainda é necessário um conhecimento estatístico para a sua utilização adequada.

Dentro de um contexto específico desse emprego estão os profissionais de Marketing, pois foram eles os principais responsáveis pela popularização das técnicas multivariadas, aplicando-as comumente em atividades como as de preferência do consumidor, satisfação de clientes, estudo de imagem e comportamento de compra. Quando se aborda as técnicas multivariadas, fala-se em técnicas que não se refere a uma dimensão da análise de dados, mas sim a uma gama de cruzamentos entre variáveis dependentes e independentes, ou mesmo um cruzamento de variáveis envolvendo informações de vários tipos, oferecendo assim ao pesquisador outras dimensões, mais ricas que a normalmente encontrada em uma abordagem univariada. (NIEDERAUER 03)

Análise por componentes Principais é o termo de uma técnica multivariada, cujo propósito principal é a redução de dados e sumarização, ou seja, a partir de uma grande massa de dados e variáveis, a obtenção de um número de combinações lineares

(componentes principais) de um conjunto dessas variáveis que tenham o máximo de informação possível que originalmente eram da grande massa original.

Ela nada mais é que cálculo dos autovalores e correspondentes autovetores de uma matriz de variâncias-covariâncias ou de uma matriz de coeficientes de correlação entre variáveis. Quando as variáveis apresentam escalas diferentes é necessário empregar alguma técnica, transformação Z, por exemplo, para a padronização. Isso se deve, pois não é possível através dessa técnica a avaliação de diferentes medidas.

A técnica de **ACP** é a mais utilizada. Consistem numa transformação linear das "*l*" variáveis originais em "*k*" novas variáveis, de tal modo que a primeira nova variável resultante seja responsável pela maior explicação da variabilidade existente no conjunto de dados, a segunda pela variação possível restante, e assim por diante até que toda a variação do conjunto tenha sido explicada. A **ACP** tem como primeiro objetivo mensurar as inter-relações entre variáveis e inicia-se com a obtenção de autovalores (*eigenvalues*) e autovetores (*eigenvectors*) a partir de uma matriz de variâncias ou covariâncias ou de correlações. (JOHNSON 1998)

A ACP busca não somente a descoberta de dimensões subjacentes, mas também a redução das variáveis originais da base de dados.

2.2 – Embasamento teórico da ACP

Para proceder a ACP, após a seleção das variáveis, procede-se uma análise descritiva dessas variáveis de forma a possibilitar inicialmente uma visualização do comportamento univariado dessas variáveis podendo ser:

- a) Matriz de médias e desvio padrão
- b) Gráfico tipo *Box Plot* ou diagrama de caixa.
 - a. Servirá para identificar isoladamente variáveis com possíveis pontos discrepantes.
- c) Gráfico da Matriz de Correlação.

- a. Servirá para identificar precisamente os possíveis pontos discrepantes da amostra, pois dá a possibilidade de analisar a variável em mais de uma dimensão.
- b. Finalizadas as análises prévias, poderão ser evidenciados valores discrepantes no conjunto de variáveis em questão.
- d) Matriz de Correlação - É uma matriz quadrada, simétrica, cuja diagonal é formada pela unidade, pois se trata da correlação da variável com ela mesma, e em cada interseção linha (i) coluna (j) a correlação das variáveis X_i e X_j .
- e) Matriz de Autovalores.
- f) Gráfico de Autovalores.
- g) Matriz de Cargas Fatoriais.
- h) Matriz de Coeficientes dos Escores Fatoriais.
- i) Matriz dos Escores Fatoriais do indivíduo.
- j) Gráficos 2D e 3D
- k) Gráficos com dispersão de variáveis e inclusão da variável suplementar.

2.2.1. REDUÇÃO DA DIMENSIONALIDADE DOS DADOS

Supondo-se ter uma tabela com p-espécies biológicas identificadas e contadas em n-observações, este arranjo constitui uma matriz de dados denominada D. Imagine que foram pesquisadas três espécies em dez amostras; então a matriz D será de ordem 10 por 3.

Adotando, adiante, a notação proposta por Lebart e Fénelon (1971: 196), em que d_{oj} representará o vetor coluna com n componentes, n observações da variável j, enquanto que d_{iv} será o vetor coluna com p componentes, as p variáveis da observação i.

Agora a matriz de dados D pode ser representada da seguinte forma:

$$D = (d_{o1}, d_{o2}, \dots, d_{oj}, \dots, d_{op})$$

e após a transposição

$$D' = (d_{1v}, d_{2v}, \dots, d_{iv}, \dots, d_{nv}).$$

Ao trabalhar em \mathcal{R}^p tem-se uma nuvem de n pontos ($n = 10$) representados no sistema de p eixos coordenados ($p = 3$).

O conjunto de pontos é uma representação unidimensional ajustando uma reta que passa pela origem, de modo que ela conserve da melhor forma a estrutura do grupo e garanta uma discrepância máxima entre as imagens dos diferentes pontos.

O critério utilizado para o ajuste é aquele que faz com que as distâncias ao quadrado desde os pontos até a reta sejam mínimas (ajuste mínimo quadrático).

Ao buscar uma representação bidimensional do conjunto, o plano sobre o qual será feita a projeção dos pontos ficará definido por t e t' sendo as retas ortogonais e t' ajustados com o mesmo critério que a primeira.

Tendo estas idéias em mente, passa-se a desenvolver o tratamento formal do problema de ajustar uma nuvem de pontos num sub-espço vetorial. Continuar-se-á primeiramente referindo-se a \mathcal{R}^p e logo analisar-se-á o caso para \mathcal{R}^n .

O problema em \mathcal{R}^p

Seja u o vetor unitário dos cossenos diretores da reta t , então a projeção de um ponto d_{iv} sobre t é igual ao produto $u \cdot d_{iv}$. Sabendo que a distância ao quadrado desde o ponto d_{iv} à origem pode decompor-se no quadrado de sua projeção sobre a reta t e o quadrado da distância do ponto à reta, o problema do ajuste mínimo quadrático pode delinear-se de dois modos: encontrar um mínimo para as distâncias desde os pontos à reta ou determinar um máximo para as projeções dos pontos sobre a reta.

Tomando a segunda opção procurar-se-á u de maneira que se tenha um

máximo na seguinte expressão:

$$\sum_{i=1}^n (u' d_{iv})^2.$$

Considerando que $(u' d_{iv})^2 = u' (d_{iv} d'_{iv}) u$, a quantidade a maximizar é:

$$u' \left[\sum_{i=1}^n d_{iv} d'_{iv} \right] u = u' D' D u$$

com a restrição de que $u' u = 1$.

Denominando-se C à matriz quadrada simétrica resultado da pré-multiplicação de D por sua transposta, a equação a maximizar é a forma quadrática $u' C u$.

Introduzindo o multiplicador de Lagrange λ e estabelecendo a seguinte expressão

$$k = u' C u + \lambda (1 - u' u)$$

e, por diferenciação, obtem-se

$$dk/du = 2Cu - 2\lambda U$$

que igualada a zero, resulta:

$$Cu - \lambda u = 0$$

ou melhor

$$(C - \lambda I) u = 0.$$

A solução não trivial do problema implica que o determinante $|C - \lambda I|$ seja singular.

Pela teoria espectral para matrizes simétricas (Morettin, 1975) sabe-se que a solução de $|C - \lambda I| = 0$, é um polinômio de ordem p em λ , tendo p raízes $\{\lambda_1, \lambda_2, \dots, \lambda_p\}$ ordenadas por seus valores em forma decrescente. Tais raízes são conhecidas como os autovalores da matriz C e o conjunto dos autovalores é o espectro de C .

Associado a cada autovalor, existe um vetor unitário, convencionalmente normalizado, que se denomina autovetor e satisfaz a relação

$$Cu_m = \lambda_m u_m \quad (m = 1, 2, \dots, p),$$

constituindo as p soluções.

É claro, que o autovetor u que soluciona o problema anteriormente proposto, é aquele que corresponde ao maior autovalor de C , isto é:

$$Cu_1 = \lambda_1 u_1.$$

Todavia, a teoria espectral indica que em matrizes simétricas reais, se dois autovalores são diferentes, seus autovetores são ortogonais. Ou seja, se $\lambda_k \neq \lambda_l$ o produto entre os correspondentes autovetores é nulo: $u_k u_l = 0$.

Isto permite dizer que t' terá como cosenos diretores os componentes do autovetor unitário u_2 associado ao segundo maior autovalor λ_2 .

Nos parágrafos anteriores foi suposto tacitamente que os p autovalores de $C(p \times p)$ eram todos diferentes. Na realidade, pode ocorrer que dois ou mais autovalores sejam iguais, e em tal caso são conhecidos como autovalores múltiplos. Adicionalmente uma ou mais raízes podem ser nulas, o que implica que o posto r da matriz C é menor que p , existindo em consequência $(p - r)$ vetores linearmente dependentes em C .

A respeito da matriz a diagonalizar C , obtida como o menor produto momento da matriz de dados, na maioria das análises multivariadas, representa a matriz de covariâncias ou a de correlações entre as variáveis. Para o cálculo delas, parte-se

respectivamente de uma matriz de desvio da média (Y) ou de uma matriz de dados padronizados (Z): agora, a matriz de covariâncias ou correlações simplesmente será obtida pelas seguintes relações:

$$C = (Y'Y) / n = S$$

$$C = (Z'Z) / n = R.$$

A seguir, denotar-se-á com a letra S a matriz de covariâncias e com R a de correlações.

O problema para \mathcal{R}^n

Se p pontos variáveis estão representados num espaço n-dimensional e deseja-se ajusta-los a uma reta v passando pela origem, expressa por um vetor unitário w de cosenos diretores, então a projeção um ponto d_{oj} será igual ao produto $w'd_{oj}$.

Maximizando a soma dos quadrados das projeções com a restrição de que $w'w = 1$, ter-se-á um ajuste mínimo quadrático para a reta. Procurar-se-á, então, que o somatória em j seja um máximo:

$$\sum_{j=1}^p (w'd_{oj})^2.$$

Desenvolvendo-se o valor quadrático como precedentemente, vê-se que o valor a maximizar é:

$$w' \left[\sum_{j=1}^p d_{ij} d'_{iv} \right] w = w'D'Dw$$

Também sabe-se que w é o autovetor da matriz DD' associado ao maior autovalor μ da mesma.

Por outra parte, o melhor sub-espço k -dimensional para projetar a nuvem de pontos ficará definido, então, pelos autovetores normalizados correspondentes aos k maiores autovalores de DD' .

Como anteriormente, pode-se referir à pós-multiplicação de D pela sua transposta, o maior produto momento a uma matriz de covariâncias ou de correlações, mas, neste caso, calculadas entre observações.

Podemos chamar tanto de redução como substituição o propósito da análise, ou seja, substituir um conjunto de variáveis correlacionadas, por um conjunto de variáveis novas não correlacionadas, sendo estas, combinações lineares das iniciais e estando ordenadas de maneira que suas variâncias decresçam da primeira à última.

Seja $D = d_{ij}$ ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, p$) a matriz de dados com p variáveis (x_1, x_2, \dots, x_p) e cujo posto r é igual ao número de variáveis ($r = p$); procuraremos então determinar y_1, y_2, \dots, y_p com as seguintes propriedades:

- 1) que cada y seja uma combinação das p variáveis x_i

$$y_1 = x_1 u_{11} + x_2 u_{12} + \dots + x_p u_{1p}$$

$$y_2 = x_1 u_{21} + x_2 u_{22} + \dots + x_p u_{2p}$$

$$\vdots$$

$$\vdots$$

$$y_p = x_1 u_{p1} + x_2 u_{p2} + \dots + x_p u_{pp}$$

- 2) que a soma dos quadrados dos coeficientes u_{ij} seja igual a 1

$$\sum_{i=1}^p (u_{ij})^2 = 1;$$

- 3) que as combinações lineares das variáveis estejam ordenadas por suas variâncias

$$\text{Var } y_1 > \text{Var } y_2 > \dots > \text{Var } y_p$$

- 4) que as novas variáveis y sejam não correlacionadas entre si.

A idéia principal é que as primeiras k novas variáveis y , os primeiros componentes principais, dêem conta da maior parte da variabilidade dos dados originais, permitindo deixar de computar $(p - k)$ componentes de menor importância. Obtém-se dessa forma uma redução do número de variáveis sem perda considerável de informação.

Pode-se dizer que a análise é simplesmente uma maneira de conseguir um modo distinto e talvez mais conveniente de expressar o mesmo conjunto de resultados.

Utilizando uma notação matricial, encontrar os componentes principais consiste em transformar a matriz de dados, que de modo genérico denota-se pela letra D e que possui p variáveis, em outra matriz F de variáveis hipotéticas não correlacionadas e cuja variância decai da primeira a última.

Para conseguir a transformação basta pós-multiplicar D por uma matriz ortogonal A , cujas colunas são, num primeiro tipo de solução, os autovetores normalizados que se calculam a partir da matriz menor produto momento de D .

Então

$$F_{(n \times p)} = D_{(n \times p)} A_{(p \times p)}.$$

As colunas de A estão ordenadas de modo que a primeira está constituída pelos p componentes do autovetor associado ao maior autovalor de $D'D$. A segunda coluna corresponde ao autovetor calculado desde o segundo maior autovalor, e assim

sucessivamente.

F é conhecida como a matriz de escores fatoriais, enquanto que A é a matriz de cargas fatoriais. A partir destas duas matrizes pode-se reconstituir a tabela original de dados ou aproximá-la se o número de componentes ou fatores extraídos for $k < p$.

Logo

$$D_{(n \times p)} = F_{(n \times p)} A'_{(p \times p)}$$

ou

$$D_{(n \times p)} \cong F_{(n \times k)} A'_{(k \times p)}$$

Para a segunda equação pode-se escrever a diferença entre o valor aproximado e o real como uma matriz de resíduos $E_{(n \times p)}$ a qual permite completar o modelo.

O Modelo

Apresentando isto mais formalmente, o modelo fica expresso pela equação

$$D = FA' + E$$

A maneira de definir os fatores assim como as suposições, com respeito à natureza dos resíduos, permitem distinguir a análise em componentes principais daquela de fatores comuns e únicos.

O segundo tipo de análise multivariada não será tratada embora possa assinalar que na mesma, a matriz de resíduos não será necessariamente pequena como é o caso dos componentes principais. Ainda mais, ela pressupõe a não correlação dos resíduos não só com os fatores, como é o caso da análise em componentes principais, mas também entre eles próprios. Deste modo na análise em fatores comuns e únicos o produto $E'E$ será uma matriz diagonal.

Sob outro ponto de vista, na análise em componentes principais, os fatores são determinados para dar conta do máximo de variância entre todas as variáveis observadas, enquanto eles são extraídos na análise em fatores comuns e únicos para responder maximalmente sobre as intercorrelações das variáveis.

Prosseguindo com o método de ACP pode-se observar que é possível uma interpretação geométrica simples. Encontra-se uma descrição muito clara para o caso bivariado em Cassie (1963).

Por exemplo, se a matriz para diagonalizar fosse aquela de covariâncias, S , então a equação $u'Su = K$ representará um elipsóide em p dimensões. Ainda, se as p variáveis têm distribuição normal multivariada, os elipsóides construídos para diferentes distâncias do centro de distribuição representam contornos com igual densidade de probabilidade.

Os cálculos necessários para encontrar os eixos principais do elipsóide, ordenados por suas longitudes, são os mesmos que aqueles empregados para extrair os componentes principais.

Para conseguir os componentes principais, como já visto quando considerou-se o problema de ajustar uma nuvem de pontos num sub-espço vetorial, tem-se que procurar os vetores que maximizem a forma quadrática $u'Su$, com a restrição de que $u'u = 1$. Isto implica em encontrar os vetores que satisfaçam a relação $Su = \lambda u$, sendo $|S - \lambda I| = 0$.

Com a finalidade de generalizar e distinguir duas soluções para a ACP retoma-se a notação utilizada quando tratou-se da estrutura básica de uma matriz.

Por outro lado, será feito a análise partindo da matriz de desvios da média onde todo elemento por sua vez é dividido pela raiz quadrada do número de observações. Designando tal matriz com a letra Y e então a matriz de covariâncias entre variáveis será simplesmente seu menor produto momento,

$$S = y'y.$$

A matriz de dados pode ser aproximada pós-multiplicando a matriz de escores fatoriais pela transposta daquela de cargas fatoriais de modo que $Y \cong FA'$. A diferença entre o valor real e o ajustado constitui a matriz de resíduos:

$$E = Y - FA'.$$

Considerando que tal diferença é pequena e conseqüentemente que a aproximação obtida por FA' é apropriada. Agora será possível descrever a estrutura básica do produto segundo a equação

$$FA' = W_{(n \times k)} \Gamma_{(k \times k)} U'_{(k \times p)}$$

Uma primeira solução consistirá nas seguintes igualdades:

$$A = U$$

e

$$F = W\Gamma.$$

Deste modo o cálculo da matriz de variâncias dos fatores é:

$$F'F = \Gamma W'W\Gamma = \Gamma^2_k = \Lambda_k.$$

Verifica-se que o resultado é a matriz diagonal dos autovalores de $S = Y'Y$ e que os mesmos ficam ordenados em forma decrescente por suas magnitudes. Os elementos fora da diagonal, todos nulos, mostram a não correlação dos fatores, que possuem variâncias diferentes.

A matriz de covariâncias dos dados será aproximada por

$$S = Y'Y \cong AF'FA' = A\Lambda A'.$$

Na prática o processo do cálculo implica em três passos: 1) obtenção de $S = Y'Y$; 2) extração dos k primeiros autovalores, sendo que eles constituem Λ_k , e computação dos correspondentes autovetores $U_{(p \times k)} = A$; 3) determinação de F pre-multiplicando A pela matriz Y .

Numa segunda solução tem-se o seguinte modo

$$F = W$$

e

$$A = U\Gamma.$$

Por esta opção obtém-se fatores não correlacionados e estandardizados de modo que

$$F'F = W'W = I_k.$$

As variâncias da matriz de cargas fatoriais são expressas por

$$A'A = \Gamma U'U\Gamma = \Gamma^2 = \Lambda_k.$$

e S é aproximada segundo

$$S = Y'Y \cong AF'FA' = AA'.$$

O procedimento prático compreende quatro etapas que são: 1) obtenção de $S = Y'Y$; 2) computação dos k autovalores e autovetores, Λ_k e $U_{(p \times k)}$; 3) obtenção de A pós-multiplicando a matriz de autovetores pela raiz quadrada de autovalores, $A = U\Lambda^{1/2}$; 4) cálculo da matriz de escores fatoriais pela equação $F = Y\Lambda^{-1}$.

Nota-se que a primeira solução é computacionalmente mais simples, porém como os fatores têm diferentes variâncias não possibilitam a comparação direta das colunas de A . Porém a segunda opção a permite.

A correlação entre variáveis e fatores pode ser estimada dividindo-se cada fila da matriz de autovetores, previamente pós-multiplicada por $\Lambda^{1/2}$, pelos desvios padrão das variáveis.

Então, na segunda solução a correlação entre fatores e variáveis (C_{fv}) é obtida por

$$C_{fv} = D_v^{-1} A$$

onde D_v^{-1} é a matriz diagonal dos inversos dos desvios padrões das Variáveis.

No primeiro tipo de solução teremos que calcular $U\Lambda^{1/2}$ ou então empregar a seguinte equação

$$C_{fv} = D_v^{-1} \Lambda \Lambda^{-1/2}$$

Os Problemas de Escala

É conhecido que, pelo caráter espectral da estrutura da matriz para

diagonalizar, a variância das variáveis fica repartida através de todos os fatores. Assim, o primeiro será uma média ponderada das variáveis presentes. Ao trabalhar partindo de S, sem transformação de dados, tem-se geralmente um primeiro componente responsável pela maior parte da variabilidade, mas que reflete principalmente a ordem de grandeza das variáveis.

Uma das idéias mais aceitas para resolver o problema é efetuar a análise partindo de R (matriz de correlações) mas agora cada variável terá a mesma influência ou igual peso.

Os componentes extraídos a partir de S ou R não serão iguais e por sua vez não é possível passar de uma solução para outra por algum coeficiente simples

Muitas das aplicações incluem o cálculo partindo de R como por exemplo ao trabalhar com variáveis medidas em unidades diferentes (m, °C, g, %, etc.). No caso contrário, ou seja quando as variáveis são razoavelmente comensuráveis, é preferível o uso da matriz de covariâncias principalmente por algumas propriedades como, por exemplo, a teoria da amostragem dos componentes disponíveis para S. Que, segundo demonstrou Anderson (1958) quando a mesma pode ser aplicada à matriz de correlações, é mais simples para S do que para R.

Adicionalmente existem outras desvantagens ao se trabalhar com R. Não é possível aplicar testes clássicos de significância ou calcular intervalos de confiança para os autovalores.

Transformação de Resultados

Anteriormente viu-se que na análise em componentes principais o primeiro fator é colocado de maneira que a variância expressa pelo mesmo constitui um máximo. O segundo fator responde por um máximo de variância remanescente, ficando num ângulo reto em relação ao primeiro. O mesmo acontece com os fatores restantes sendo todos ortogonais.

Pode-se dizer então que a análise além de determinar a dimensionalidade dos dados, proporciona uma única maneira de posicionar o conjunto dos fatores.

Ao transformar linearmente os vetores que definem um espaço, o novo conjunto obtido conservará as mesmas propriedades; poderemos então achar infinitas soluções para o modelo. Assim, tendo-se uma matriz de transformação $T_{(k \times k)}$ não singular, as pós-multiplicações da matriz de escores fatoriais pela transposta de T e da matriz de cargas fatoriais pela inversa de T , dão como resultado F^* e $A^{*'}$ que igualmente satisfazem o modelo

$$Y = FA' + E = FT'A'T^{-1} + E = FT'T'^{-1}A' + E = F^*A^{*'} + E.$$

Qualquer par de fatores equivalentes em F e F^* serão indistinguíveis e responsáveis pela mesma quantidade de variância dos dados.

Sendo T ortonormal, $T' T = 1$, a transformação de F para F^* corresponde à rotação rígida dos eixos não modificando o ângulo de separação entre eles. Em caso contrário a solução é chamada oblíqua e os fatores resultantes são correlacionados.

A rotação dos eixos não modifica a configuração dos pontos variáveis e as cargas fatoriais sobre os novos eixos obtidos descreve o posicionamento deles com a mesma fidelidade que antes da transformação.

As infinitas soluções do modelo, anteriores, correspondem ao número de

posições que podem adotar os eixos sem modificar a validade da análise.

O objetivo de efetuar a rotação dos fatores após se obter a solução em componentes principais é conseguir uma matriz de cargas fatoriais que se possa interpretar com maior facilidade, isto é, que permita identificar certos fatores tendo como base as altas saturações das cargas de algumas variáveis.

Foi em meados do século passado, que Thurstone (1949) criou o conceito de estrutura simples. As suas idéias permitiram achar soluções apropriadas para diferentes tipos de rotações, inicialmente na base de métodos gráficos e atualmente mediante procedimentos analíticos.

A hipótese de estrutura simples pode ser sintetizada por alguns critérios básicos:

- 1) cada fator terá um número de valores nulos, equivalentes à quantidade de fatores retidos na análise;
- 2) cada variável possuirá no mínimo uma saturação nula para um fator;
- 3) para todo par de fatores haverá três possibilidades que são: variáveis com cargas altas em um deles e baixa no outro; variáveis com cargas pequenas em ambos e finalmente poucas que poderão ter saturações importantes nos dois.

A Matriz de Escores Fatoriais

Para computar a matriz de escores fatoriais observa-se na literatura sobre o assunto diferentes procedimentos, mas nem todos são inteiramente satisfatórios.

Num trabalho de Trochinczyk e Chayes (1978) se destaca que, como resultado de transformações inconsistentes nos dados, podem ser verificadas correlações

entre os escores.

Tem-se matrizes com escores fatoriais correlacionados entre si, nos seguintes casos:

- 1) quando os componentes são extraídos a partir da matriz de correlações, R , e a matriz de escores fatoriais calculada com dados não padronizados;
- 2) quando os componentes são calculados desde a matriz de covariância, S , e os escores computados segundo a relação $F=P'Z$, onde P é a matriz de autovetores normalizados e Z a matriz de dados padronizados.

Apresenta-se em seguida algumas das soluções válidas para o cálculo da matriz de escores fatoriais quando:

- a) os componentes são calculados a partir da matriz de correlações, R :

$$1) F = A'R^{-1} Z$$

$$2) F = Q'Z;$$

- b) os componentes são calculados a partir da matriz de covariâncias, S :

$$1)F = B'S^{-1}X$$

$$2)F = P'X.$$

A e B são matrizes de cargas fatoriais, Q e P matrizes de autovetores normalizados, Z a matriz de dados padronizados e X a matriz de dados iniciais.

A análise por componentes principais não devem ser interpretadas como sinônimo de análise fatorial. Esta que pressupõe que as relações existentes dentro de um conjunto de “ m ” variáveis seja o reflexo das correlações de cada uma dessas variáveis com “ p ” fatores, mutuamente não correlacionáveis entre si, sendo “ p ” menor que “ m ”.

Visto os aspectos teóricos que envolvem a análise por componentes principais essenciais para o desenvolvimento do módulo ACP para o SESTATNET é oportuno uma abordagem sobre a Inteligência artificial e a Educação.

3 – A Inteligência Artificial e a Educação.

Desenvolver sistemas especialistas a fim de torná-lo um agente que viabilize mais e melhor a aprendizagem não constitui um avanço para a educação. Mesmo os melhores sistemas especialistas produzidos não têm o perfil necessário de um educador. Então entender, mesmo que superficialmente, como é o processo educacional em suas tendências ditas tradicionais e modernas, é muito importante. O subsídio e enfoque pedagógico que o SESTATNET agrega juntamente com suas características de inteligência artificial e sistemas especialistas é que faz dele um sistema diferente. Como apoio ao ensino tradicional ou a distância o SESTATNET é uma ferramenta de disponibiliza o conhecimento técnico e a forma pedagógica tão importante na aprendizagem e na aquisição do conhecimento.

3.1 - O Construtivismo e a Educação Tradicional

Os conceitos educacionais apresentados pelo **SESTATNET** e conseqüentemente pelo módulo **ACP** têm sua fundamentação voltada para as teorias construtivistas. Sendo um sistema especialista como todo o **SESTATNET** apresenta seu conteúdo de especialista de uma forma a propiciar a aprendizagem de uma forma dinâmica e construtivista.

Aos recursos agregados nessa versão os conceitos de educação a distância (EaD) também se aplicaram ao SESTATNET que propicia a relação aluno-professor dimensões extra classe.

Ao se abordar conceitos tão amplos e dignos de profundos estudos como os relacionados à educação não se pretende criar definições, mas sim uma base para o melhor entendimento a cerca de toda a dinâmica que envolve esses assuntos. Adiante se tenta mostrar o quão amplo pode ser assunto.

Quando um paciente vai ao médico e ele manda tomar remédios e dá a receita, alguém o questiona? A maioria das pessoas dirá que não, pois, se existe uma classe de profissionais que é especializada em conhecer os medicamentos e substâncias corretas, os demais estão liberados de preocupar-se com esse funcionamento e podem simplesmente fazer o que o médico manda.

Ninguém que frequente o consultório em busca do conhecimento do médico acredita que precise pensar muito a respeito. O médico manda! Os que não são médicos obedecem.

Sem querer fazer referência aos médicos somente, isso pode aplicar-se a quase tudo: Os economistas, engenheiros, urbanistas, políticos, todos decidem como será a vida da população e quase nunca ou nunca são questionados.

Assim estava também imagem do professor no Ensino Tradicional.

O professor era o único detentor do conhecimento e por isso cabia-lhe a missão de transmitir aos seus alunos a resposta correta ou o caminho para encontrá-la, assumindo papel de transmissor do conhecimento.

Era um profissional dotado de muita autoridade e praticamente inacessível, como era possível verificar pelo seu comportamento em sala de aula e pelo comportamento de complacência, aceitação e apatia que era exigido dos alunos.

E muitos séculos de educação concebida desta maneira se passaram, até que o filósofo Jean Piaget, após observar crianças, chegou a uma conclusão que viria mudar os rumos da educação.

Piaget descobriu que as crianças não pensam como adultos e que têm sua própria ordem e sua própria lógica, e devem ser respeitadas desta maneira. **(PILETTI 94)**

Nascia desta afirmação os pilares de um método conhecido como Construtivismo, que passou a questionar e até a considerar incorreta a pedagogia tradicional.

O método denominado tradicional considerava os alunos como seres a serem “preenchidos” com o conhecimento dos mestres e a partir de Piaget, investiu-se na idéia de que os alunos são como criadores e construtores de seus próprios conhecimentos e que a melhor forma de fazer isso é permitir-lhes criar e testar suas teorias sobre as coisas a sua volta, concluindo-se assim que todo o conhecimento vem, na verdade, da interação entre o

sujeito e o mundo que o cerca ou da “assimilação da realidade”, nas palavras de Piaget. **(PILETTI 94)**

Construtivismo afirma que para que haja aprendizagem o sujeito do aprendizado deve fazer relação constante entre o que se lhe apresenta como novo e o que já conhece, assim, estabelece relações e vai construindo seu conhecimento.

O professor tem aí o papel de facilitador dessa aprendizagem e não mais o papel de transmissor de conhecimento, pois a transmissão já não satisfaz as necessidades do aluno, mesmo que ela seja feita de forma didaticamente competente e eficiente.

Então, o método construtivista difere do tradicional porque o tradicional “valorizava o desenvolvimento intelectual a partir da ação dos indivíduos e da força do meio em que conviviam”, explica Márcia Moreira, professora de Psicologia Educacional da Pontifícia Universidade Católica (PUC) de Belo Horizonte, e o construtivismo atribui um papel ativo ao indivíduo, sob a influência do meio. “É a pessoa que constrói o seu próprio conhecimento”. **(PILETTI 94)**

O Construtivismo tem a finalidade principal de propiciar ao aluno condições para que resolva por si próprio seus problemas, através dessa interação com o meio que o cerca e com a ajuda do facilitador, papel desempenhado pelo professor, não aceitando as tradicionais idéias de formar o aluno de acordo com modelos prévios e condução às respostas esperadas. No quadro 1 apresenta-se um comparativo entre os métodos tradicional e o construtivista.

Quadro 1 - Comparativo entre as salas de aula Tradicional e Construtivista, segundo (Brooks e Brooks, 1995)

Sala de aula Tradicional	Sala de aula Construtivista
O conteúdo é apresentado das partes para o todo, com ênfase nas habilidades básicas	O conteúdo é apresentado do todo para as partes, com ênfase nos conceitos gerais
O seguimento rigoroso do conteúdo pré-estabelecido é altamente valorizado	A Busca pelas questões levantadas pelos alunos é altamente valorizada

As atividades baseiam-se fundamentalmente em livros texto e de exercícios.	As atividades baseiam-se em fontes primárias de dados e materiais manipuláveis.
Os estudantes são vistos como "tábuas rasas" sobre as quais a informação é impressa.	Os estudantes são vistos como pensadores com teorias emergentes sobre o mundo
Os professores geralmente comportam-se de uma maneira didaticamente adequada, disseminando informações aos estudantes.	Os professores geralmente comportam-se de maneira interativa, mediando o ambiente para estudantes.
O professor busca as respostas corretas para validar a aprendizagem	O professor busca os pontos de vista dos estudantes para entender seus conceitos presentes para uso nas lições subseqüentes.
Avaliação da aprendizagem é vista como separada do ensino e ocorre, quase que totalmente através de testes	Avaliação da aprendizagem está interligada ao ensino e ocorre através da observação do professor sobre o trabalho dos estudantes
Estudantes trabalham fundamentalmente sozinhos	Estudantes trabalham fundamentalmente em grupos

A concepção do SESTATNET é centrada no construtivismo, visto que a grande maioria de suas características aplica-se perfeitamente no modelo de ensino que ele se propõe.

Uma das várias características que fica fortemente destacada ligando as teorias construtivistas e o SESTATNET é a capacidade do indivíduo (aluno) ser ativo na construção do seu conhecimento estatístico.

3.2 – A Educação a distância

A educação a distância ou o ensino a distância não são conceitos unânimes. Ambos são comumente apresentados como formas ou processos de ensino-aprendizagem nos quais os elementos desse processo (aluno/professor) estão separados de alguma forma

fisicamente. As formas de comunicação entre eles são as mais diversas, tais como: rádio, televisão, cartas, fitas de vídeo, fitas de áudio, internet, fax, CD-ROM, telefone e outras tecnologias semelhantes. Alguns preferem o termo educação a distância por entenderem que é um conceito mais amplo do que ensino a distância.

Para entender Educação a Distância é importante lembrar que a Ead pressupõe um sistema de transmissão e estratégias pedagógicas adequadas às diferentes tecnologias utilizadas. Essas estratégias são as escolhas das formas instrucionais que terão a finalidade de produzir o aprendizado do aluno e na escolha dessas estratégias a atenção deve ser voltada não somente para a ementa, o conteúdo, mas também à escolha de um corpo docente preparado e habilitado para lidar com as novas tecnologias que se apresentam modernamente e muito mais, também para um eficiente método de suporte ao aluno, como a presença de um tutor permanente, que pode ser o docente que ministrou as instruções iniciais ou outro, com capacitação para tal, também chamado de tutor, que apoiará e auxiliará o aluno em suas pesquisas e produções, proporcionando-lhe condições favoráveis para seu aprendizado.

O aluno, na Ead, deve ter um perfil que possibilite ter condições de um aprendizado autônomo, pois nessa modalidade de ensino terá uma instrução sobre o conteúdo de interesse, mas não lhe será dado o lugar de pessoa a ser educada por outro, de *tábula rasa*, como frequentemente acontecia no ensino tradicional mas esperar-se-á dele que seja autônomo, investigativo, disciplinado e até mesmo que resgate a curiosidade, tão natural no aprendizado infantil mas que aos poucos se vai perdendo com o transcorrer da vida acadêmica.

Pois é justamente da curiosidade e disciplina, aliadas ao material didático e instruções recebidas de que o aluno será capacitado no Ead, ou seja, a autonomia é de extrema relevância, já que os materiais são oferecidos e apresentados ao aluno, não impostos a ele.

A EaD caracteriza-se pela possibilidade de estabelecer comunicação de múltiplas vias e seus referências são fundamentados pelos pilares da Educação publicados pela UNESCO: *aprender a conhecer, aprender a fazer, aprender a ser e aprender ao longo da*

vida e assim a educação deixa de ser concebida como transferência de informações e passa a ser norteadada pela contextualização de conhecimentos úteis ao aluno.

O desenvolvimento da Ead pode ser basicamente descrito em três etapas, ou gerações:

- A primeira geração tem como característica o material impresso e foi iniciado no século XIX.

- A segunda geração contou com os recursos de rádio e televisão, aulas expositivas, fitas de vídeo. Nessa fase, o Instituto Universal Brasileiro desempenhou um papel significativo na EaD no Brasil e está há mais de 60 anos nessa modalidade educativa.

- A terceira geração é marcada pelos recursos de WEB, como teleconferência, chat, fóruns, correio eletrônico e plataformas de ambientes virtuais que possibilitam a interação entre alunos e professores.

A EaD, em sua concepção principal que é a de proporcionar condições de aprendizado a alunos e professores separados fisicamente, é conhecida desde o século XIX e tem como um de seus primeiros marcos uma publicação feita no jornal *Gazeta de Boston*, no dia 20 de março de 1728, pelo taquigrafista Cauleb Phillips: *“Toda pessoa da região, desejosa de aprender esta arte, pode receber em sua casa várias lições semanalmente e ser perfeitamente instruída, como as pessoas que vivem em Boston”*, mas apesar de ser conhecida desde então, não fez parte das preocupações pedagógicas a não ser nas últimas décadas no século XX.

A educação a distância tornou-se mais visível e teve maior atenção quando tornou-se desejável e até mesmo inevitável a formação profissional e até cultural de pessoas que não tinham condições de frequentar um ambiente formal de educação e a partir da primeira Grande Guerra Mundial surgiram novas iniciativas de ensino a distância pois houve um considerável aumento da demanda social por educação.

A modernização de serviços como o de correio e a agilidade dos meios de transporte, além da aplicabilidade de novas tecnologias nos meios de comunicação influenciaram bastante essa modalidade de ensino e durante as primeiras décadas do século XX, a União Soviética e a França já organizavam sistemas de ensino por correspondência.

Desde então o rádio começou a ser utilizado para fins de educação a distância pois alcançava meios formais e informais e foi largamente utilizado na América Latina nos programas de Educação a Distância e a partir de 1960 e 1970, apesar do material escrito ainda prevalecer, incorporou-se também videocassetes e fitas e como não poderia deixar de ser, o meio de comunicação que se tornaria na próxima década um dos mais importantes meios de divulgação e instrução para a EaD: o computador.

Atualmente é utilizada toda tecnologia de multimeios conhecida, possibilitando que quase todo tipo de aprendizado se renda ao ensino não presencial e confirmando a EaD como uma modalidade sólida de ensino-aprendizagem que acompanha todo o desenvolvimento experimentado pela humanidade em relação a tecnologia e o que antes era utilizado apenas como recurso para superação de deficiências educacionais, aperfeiçoamentos ou qualificação profissional, hoje é cada vez mais encontrado como programas de ensino que complementam formas tradicionais de educação.

Existem dois meios de ensinar através de e-learning, o síncrono e o assíncrono. O síncrono caracteriza-se pela comunicação simultânea entre professor e aluno, que pode se dar através de diversos meios como chats, video conferência e telefone e o assíncrono é aquele em que o professor e aluno não estão ao mesmo tempo em sala de aula e a comunicação pode ser feita por meio de emails ou fóruns, por exemplo.

Além da diferença temporal há também outras diferenças entre síncrono e assíncrono. Enquanto esse tem baixo custo por número de alunos aquele assemelha-se ao ensino tradicional pois o professor ministrará aulas ao vivo e o aluno, via WEB, assistirá a aula onde estiver e terá a oportunidade de fazer suas perguntas em momentos de aulas ou ainda de atividades proporcionadas via chat, porém a grande diferença entre os dois métodos de e-learning é sem dúvida a maneira de utilização do tempo pois na atividade assíncrona o tempo é mais flexível, permitindo ao aluno pensar, pesquisar e estudar antes de fazer uma atividade e encaminhá-la ao professor ou ao tutor que lhe dará o suporte enquanto pelo meio síncrono o tempo é mais rígido e exige resposta rápidas.

É claro que o EaD também exigiu dos docentes e educadores novos esquemas e novas concepções acerca do ensino, obrigando-os a uma maior criatividade, utilização e pesquisa de recursos tecnológicos, elaboração de aulas com conteúdos mais condensados e

além de tudo isso acostumar-se com o que para alguns ainda é fundamental mais quase impossível na EaD, a não ser via web cam, que é o “*olho no olho*”, tão comum na educação tradicional e extremamente valorizado pela cultura brasileira e não encontrado com essa mesma intensidade na cultura européia, de onde muitos modelos de EaD foram importados.

Sendo assim fica cada vez mais claro e evidente que na sociedade atual não é mais possível pensar em educação como repasse de conhecimentos e é necessário pensar e repensar novas formas de educação que se apresentam e exigem que o processo ensino-aprendizagem se modifique frente a globalização e necessidades modernas encontradas.

É imprescindível que o processo ensino-aprendizagem moderno não ignore as habilidades e competências do aprendiz assim como possibilite que ele possa pesquisar e estudar de acordo com suas possibilidades e ao educador, por sua vez, possibilite tornar o ambiente dinâmico, flexível, promover comunicação bilateral e contextualizar o assunto abordado de acordo com o interesse do aluno, dando-lhe o papel mais importante no processo e permitindo assim que o aluno seja protagonista de seu processo de aprendizagem, independente de seu tempo para dedicação aos estudos ou mesmo de sua condição social ou ocupação.

3.3 - Inteligência Artificial.

O termo “inteligência artificial” nasceu em 1956 no famoso encontro de Dartmouth. Dentre os presentes a este encontro incluíam-se Allen Newell, Herbert Simon, Marvin Minsky, Oliver Selfridge e John McCarthy. No final dos anos 50 e início dos anos 60, os cientistas Newell, Simon, e J. C. Shaw introduziram o processamento simbólico. Ao invés de construir sistemas baseados em números, eles tentaram construir sistemas que manipulassem símbolos. **(CHAIBEN 03)**

Desde então, as diferentes correntes de pensamento em IA têm estudado formas de estabelecer comportamentos “inteligentes” nas máquinas. Portanto, o grande desafio das pesquisas em IA, desde a sua criação, pode ser sintetizado com a indagação feita por

Minsky em seu livro “Semantic Information Processing”, “*Como fazer a máquinas compreenderem as coisas ?*” (MINSKY 68).

Um sistema IA não é capaz somente de armazenamento e manipulação de dados, mas também da aquisição, representação, e manipulação de conhecimento.

Esta manipulação inclui a capacidade de deduzir ou inferir novos conhecimentos - novas relações sobre fatos e conceitos - a partir do conhecimento existente e utilizar métodos de representação e manipulação para resolver problemas complexos que são freqüentemente não-quantitativos por natureza. Uma das idéias mais úteis que emergiram das pesquisas em IA, é que fatos e regras - conhecimento declarativo - podem ser representados separadamente dos algoritmos de decisão - conhecimento procedimental ou procedural. Isto teve um efeito profundo tanto na maneira dos cientistas abordarem os problemas, quanto nas técnicas de engenharia utilizadas para produzir sistemas inteligentes. Adotando um procedimento particular - máquina de inferência - o desenvolvimento de um sistema IA é reduzido à obtenção e codificação de regras e fatos que sejam suficientes para um determinado domínio do problema.

Portanto, as questões principais a serem contornadas pelo projetista de um sistema IA são: aquisição, representação e manipulação de conhecimento e, geralmente, uma estratégia de controle ou máquina de inferência que determina os itens de conhecimento a serem acessados, as deduções a serem feitas, e a ordem dos passos a serem usados. A figura 2 retrata estas questões, mostrando a inter-relação entre os componentes de um sistema clássico de IA (SCHUTZER 87).

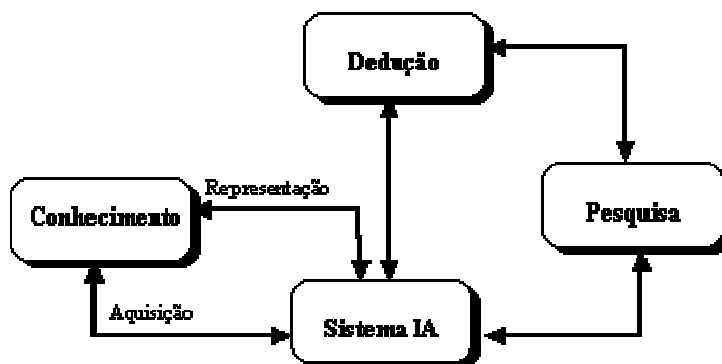


Figura 2 - Uma visão conceitual dos sistemas de Inteligência Artificial.
(SCHUTZER 87)

O primeiro sistema especialista baseado em conhecimento foi escrito em 1967. Chamado DENDRAL (BUCHANAN 78), podia prever as estruturas de compostos químicos desconhecidos baseados em análises de rotinas. Posteriormente, sistemas especialistas baseados em regras mais sofisticados foram desenvolvidos, notavelmente o programa MYCIN (SHORTLIFFE 76), que utiliza regras derivadas do domínio médico para raciocinar (deduzir) a partir de uma lista de sintomas de alguma doença em particular.

Muitos pesquisadores acreditam que IA é uma tecnologia chave para o software do futuro. As pesquisas em IA estão relacionadas com áreas de aplicação que envolve o raciocínio humano, tentando imitá-lo e realizando inferências. Estas áreas de aplicação que geralmente são incluídas nas definições de IA incluem, entre outras. (SAVORY 88)

- Sistemas Especialistas ou Sistemas Baseados em Conhecimento.
- Sistemas Inteligentes/Aprendizagem.
- Compreensão/Tradução de Linguagem Natural
- Compreensão/Geração de voz
- Análise de imagem e cena em tempo real
- Programação Automática.

Portanto, pode-se afirmar que o campo de IA tem como objetivo, o contínuo aumento da “inteligência” do computador, pesquisando, para isto, também os fenômenos da inteligência natural. Para este fim, IA é definida aqui como sendo uma coleção de técnicas suportadas por computador emulando algumas capacidades dos seres humanos. Esta coleção inclui (**SAVORY 88**):

- Resolução de problemas
- Compreensão de Linguagem Natural
- Visão e Robótica
- Sistemas Especialistas e Aquisição de Conhecimento
- Metodologias de Representação de Conhecimento

A esperança de grandes descobertas futuras em IA depende de vários fatores, tais como o crescimento do número de cientistas envolvidos nas pesquisas e avanços principalmente nas áreas da ciência da computação (incluindo processamento paralelo) e da ciência cognitiva.

A inteligência artificial pode ser definida como um campo científico preocupado com a criação de sistemas computadorizados que podem atingir níveis humanos de raciocínio. Mais precisamente, a IA é o ramo da informática que enfoca o desenvolvimento de programas de computadores capazes de desempenhar tarefas normalmente associadas ao comportamento humano inteligente.

Isso é obtido através de uma coleção de técnicas sustentadas por computadores que competem com algumas das habilidades naturais dos seres humanos. Exemplos disso são as representações do conhecimento, as capacidades de inferência, a resolução de problemas, o engajamento num diálogo e a compreensão da linguagem natural, bem como o reconhecimento e síntese da fala, a visão computadorizada e a robótica dentre outras tantas. A inteligência artificial pode também ser definida com um ramo da informática cujo objetivo é equipar as máquinas com raciocínio e capacidades perceptivas.

As características rudimentares da IA são: manipulação de símbolos em vez de números, a capacidade de fazer inferências e deduções a partir da informação disponível;

aplicar o conhecimento na resolução de problemas e a utilização do seu conhecimento e suas regras associadas para limitar o crescimento exponencial que ocorre em situações complexas no mundo real. (CHORAFAS-88)

**Quadro 2 - Comparação entre a Inteligência Artificial e a Inteligência Natural.
(CHORAFAS-88)**

Inteligência Artificial	Inteligência Natural
Igual	Igual
Cresce com a experiência	Cresce com a experiência
Torna-se obsoleta se não for utilizada	Torna-se obsoleta se não for utilizada
Positivo	Negativo
Sobrevive ao ciclo de vida do perito	Desaparece após a morte do perito.
É consistente	É irregular
Acumula-se permanentemente	É efêmera, oscilante
É bem documentada	É mal documentada
É facilmente duplicada	É difícil de transmitir
Seu custo é razoável	Custa caro
Negativo	Positivo
Atualmente não é criativa	É criativa
É ainda alheia à cultura atual	É parte da cultura atual

A utilização de computadores na educação tem uma longa história. As aplicações educacionais utilizando-se desta tecnologia são desenvolvidas desde os anos 60. Muitas são classificadas como Instrução Assistida por Computador (“CAI”, do inglês “Computer-Assisted Instruction”) e utilizam o paradigma da instrução programada, cujos métodos educacionais apresentam uma forma expositiva *centrada no professor*, ou seja, primeiramente o estudante deve compreender a lição dada pelo professor para posteriormente responder alguma questão e, com isso, reforçar a sua compreensão. (CHAIBEN 03). Esta abordagem centrada no professor foi fortemente influenciada pela Teoria Comportamentalista de Skinner (PARK 87) e tem suas raízes na psicologia do condicionamento operante (ou seja, os princípios nos quais as mudanças de comportamento - consideradas aqui como sinônimo de ensino - podem ser mais facilmente provocadas através de “programações de reforço”, isto é, compensando o comportamento desejado em dados momentos) (RICHMOND 75).

Segundo (CHAIBEN 03), existe uma crescente tendência de incorporar princípios e estratégias cognitivas no processo de desenvolvimento destes sistemas CAI, a medida que as perspectivas teóricas dos psicólogos educacionais tendem a migrar para a psicologia cognitiva.

Com a evolução das técnicas de Inteligência Artificial (IA) e das pesquisas no campo das ciências cognitivas, aumentou-se o grau de “inteligência” dos sistemas educacionais e antigas dificuldades estão sendo aos poucos superadas (CHAIBEN 03). Chamados ICAI para “**inteligente**” CAI, estes sistemas apresentam uma rica representação de seu domínio permitindo utilizar seus conhecimentos de maneira não diretamente explicitados pelo projetista.

Uma das principais motivações para as pesquisas em Inteligência Artificial na Educação (“AI-ED” do inglês “**Artificial Intelligence in Education**”), é o desenvolvimento de princípios pelos quais os ambientes de aprendizagem computacionais possam ser concebidos como lugares onde os estudantes possam ter experiências de aprendizagem individualizadas, isto é, experiências que sejam fundamentais e benéficas para eles, sem importar suas diferenças individuais, experiências anteriores, ou outras situações cognitivas (AKHRAS 95).

Assim, pela modelagem do estudante, estes sistemas buscam *personalizar a instrução*, compatibilizando a apresentação com o nível de conhecimento do estudante e com o seu índice de aprendizagem (RICKEL 89). Portanto, a maioria destes sistemas apresenta métodos educacionais que proporcionam uma forma de descoberta *centrada no estudante*, e os diálogos tutoriais são basicamente determinados pelo conhecimento conceitual e pelo comportamento de aprendizagem do estudante pressuposto pelo professor (PARK 87).

O SESTATNET não pressupõe um modelo de estudante. É um sistema aberto quanto ao usuário, pois disponibiliza o raciocínio estatístico de análise de dados, extraído de professores do departamento de Informática e Estatística da Universidade Federal de

Santa Catarina (INE-USFC). Para disponibilizar o raciocínio estatístico o SESTATNET foi desenvolvido utilizando conceitos de inteligência artificial, sendo um sistema especialista.

3.4 - Sistemas Especialistas

São sistemas computacionais inteligentes que emulam o comportamento de especialistas humanos em algum domínio específico do conhecimento. Este faz uso intensivo do conhecimento especializado para resolver problemas ao nível de um especialista humano.

Como características podem-se citar a capacidade de aprender, resolver problemas, planejar e prever conseqüências, habilidade de prever fatos ambíguos, criatividade e o uso de linguagem natural. (**LÓPEZ-03**)

Segundo (**CHAIBEN 03**) sistemas especialistas, são aplicações da inteligência artificial, são programas de computador planejados para adquirir e disponibilizar o conhecimento operacional de um especialista humano.

São tradicionalmente vistos como sistemas de suporte à decisão pois são capazes de tomar decisões como especialistas em diversas áreas e sua estrutura reflete a maneira como o especialista humano arranja e faz inferência sobre o seu conhecimento.

Os sistemas especialistas são diferentes das aplicações típicas de que? por causa de sua arquitetura. Um dos princípios fundamentais no projeto de sistemas especialistas é a separação do conhecimento de domínio (por exemplo, medicina ou geologia) dos programas que “raciocinam” com este conhecimento (**BUCHANAN 89**). Portanto, existe uma distinta divisão entre o componente de conhecimento do sistema e o componente de raciocínio ou máquina de inferência. A máquina de inferência é bem generalizada e usualmente poderá trabalhar com diferentes conjuntos de conhecimento (**PEPER 91**).

O componente de conhecimento e o componente de raciocínio são as chaves de qualquer sistema que reflita “inteligência”. Portanto, a única maneira destes sistemas apresentarem um “comportamento inteligente” é através de mecanismos formais para a representação do conhecimento e a utilização de técnicas de inferência. **(CHAIBEN 03)**

O sistema especialista mais frequentemente citado é o MYCIN, desenvolvido por uma equipe de médicos e especialistas em IA na Universidade de Stanford. O MYCIN contém o conhecimento dos mais destacados especialistas no campo de doenças infecciosas. Foi projetado para auxiliar no diagnóstico e tratamento de meningite (inflamação das membranas que envolvem o cérebro e a medula espinhal) e bacteriemia (infecção bacteriana no sangue). O MYCIN utiliza o tipo de raciocínio “*backward chaining*”. Dando-se um conjunto de sintomas para diagnóstico, o MYCIN utiliza seus conhecimentos para conduzir às conclusões e então recomendar o apropriado tratamento. O MYCIN foi um pioneiro entre os sistemas especialistas e representa um esforço de aproximadamente 50 homens/ano. Muito deste esforço está embutido em sua base de conhecimento.

Um modelo básico da arquitetura dos sistemas especialistas pode ser apresentado conforme na figura 3, com três componentes básicos: a base de conhecimento, a máquina de inferência, e a interface com usuário.

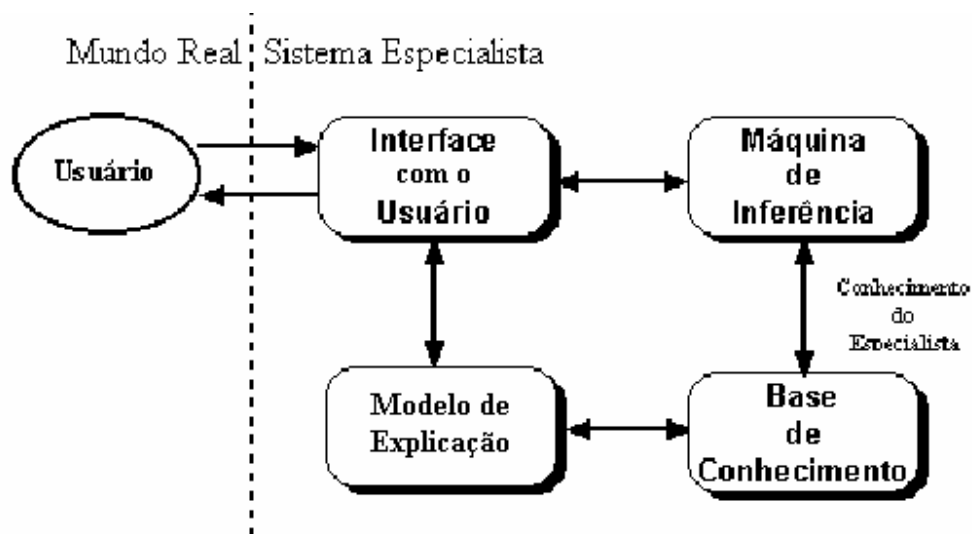


Figura 3 - Estrutura convencional de um Sistema Especialista (CHAIBEN 03)

3.4.1. - A Base de Conhecimento

A característica principal de um sistema especialista é o uso do conhecimento específico sobre um domínio de aplicação. Neste sentido, o termo “base de conhecimento ” é utilizado para significar a coleção de conhecimento do domínio, ou seja, as informações especializadas e necessárias para resolver problemas de um domínio específico. Portanto, este conhecimento precisa ser organizado de uma maneira adequada para que a máquina de inferência consiga tratá-lo convenientemente.

O conhecimento em um sistema especialista consiste de *fatos e heurísticas*. Os fatos constituem as informações que estarão sempre disponíveis para serem compartilhadas e atualizadas pelo especialista do domínio, ou pelo próprio sistema. As heurísticas são regras práticas que caracterizam o nível de tomada de decisão do especialista em um domínio. Portanto, uma base de conhecimento pode ser vista como um conjunto de regras, cada qual podendo ser validada independentemente de estrutura de controle.

Um dos problemas mais sérios, e ao mesmo tempo muito comum, encontrado na implementação de sistemas especialistas, é que usualmente parece impossível fornecer um conhecimento completo sobre o qual o sistema vai operar. Portanto, o nível de desempenho de um sistema especialista está relacionado ao tamanho e a qualidade de sua base de conhecimento.

3.4.2 - A Máquina de Inferência

Segundo Minsky, “... o conhecimento é útil somente quando se pode explorá-lo para ajudar a alcançar nossos objetivos.” (MINSKY 86).

Nos sistemas especialistas, a máquina de inferência cumpre este papel, representando o meio pelo qual o conhecimento é manipulado, utilizando-se das informações armazenadas na base de conhecimento, para resolver problemas. Para isto, deve haver uma linguagem ou um formato específico no qual o conhecimento possa ser expresso para permitir o “raciocínio” e inferência. Métodos de inferência são necessários para fazer uso apropriado e eficiente dos itens em uma base de conhecimento para alcançar alguns propósitos tal como o diagnóstico de doenças.

A máquina de inferência, de certo modo, tenta imitar os tipos de raciocínio que o especialista humano emprega quando resolve um problema, ou seja, ele pode começar com uma conclusão e procurar uma evidência que a comprove, ou pode iniciar com uma evidência para chegar a uma conclusão. Em sistemas especialistas, estes dois métodos são chamados de “*backward chaining*” e “*forward chaining*”, respectivamente.

Nem todos os sistemas utilizam a mesma abordagem para a representação do seu conhecimento, portanto, a máquina de inferência deve ser projetada para trabalhar com a representação de conhecimento específica utilizada.

3.4.3. - A Interface com o Usuário

A interface com o usuário visa facilitar a comunicação entre o sistema especialista e o usuário, permitindo a interação com o sistema através da entrada de fatos e dados e através da saída em forma de perguntas, conclusões e explicações.

Muitos princípios baseados nas teorias cognitivas têm sido propostos para projetos de interface, como resultado de pesquisas na área de interação homem-máquina. Uma das considerações principais no projeto de qual quer interface homem-máquina deve ser a facilidade de uso, reduzindo ao máximo a carga cognitiva sobre o usuário.

3.5 - Sistemas Especialistas Apoiando a Educação

Do ponto de vista educacional, a maioria dos sistemas especialistas tem pouca utilidade direta, porque não foram projetados para ensinar. Isto pode ser explicado principalmente com relação a (**TROLLIP 91**):

- *ausência de qualquer estratégia educacional.*
- *incapacidade de comparar o que o estudante conhece com o conhecimento do especialista.*
- *incapacidade de determinar o que fazer quando o conhecimento do estudante difere do conhecimento do especialista.*

Entretanto, a estrutura do sistema especialista serve muito bem para ser adaptada para a construção de sistemas tutoriais, proporcionando um grande potencial para a criação de novos ambientes educacionais. O sistema GUIDON (**CLANCEY 87^a**), por exemplo, é um sistema tutorial especialista para o ensino de diagnóstico de doenças infecciosas do sangue, que foi desenvolvido a partir da base de conhecimento já formada do MYCIN, ou seja, é um sistema especialista adaptado ao ensino.

A partir do sistema GUIDON várias lições importantes puderam ser aprendidas, pois este sistema mostrou-se inadequado principalmente por causa da representação de conhecimento baseada nas regras originais do MYCIN. O GUIDON ajudou a demonstrar a necessidade de uma base cognitiva para o componente contendo o conhecimento especialista. Portanto, um sistema tutorial não necessita somente do conhecimento de seu domínio, mas também da perspectiva sobre este conhecimento que permita transmiti-lo ao estudante adequadamente (**RICKEL 89**).

Uma característica importante nos sistemas especialistas é a separação do conhecimento dos métodos gerais que são usados para manipular este conhecimento. Esta característica é relevante se a aplicação é voltada ao ensino, uma vez que cada domínio (área de aplicação) tem sua própria terminologia, relações e procedimentos.

Se os aspectos relacionados ao domínio podem ser formulados independentemente, então o desenvolvimento completo de um sistema de ensino pode ser bastante simplificado (**KEMP 92**).

Os sistemas CAI tradicionais, embutem sua pedagogia dentro do conteúdo da lição. Isto é, técnicas educacionais bem específicas são selecionadas e talhadas para um particular conteúdo da lição e o resultado imediato disto é que estas lições proporcionam uma adaptação limitada para as necessidades de cada estudante.

Por outro lado, um professor vendo as dificuldades de um aluno, pode tomar inúmeras atitudes para ajudá-lo, fazendo analogias, propondo leituras adicionais, ou simplesmente dando um tempo ao aluno. É esta *versatilidade* humana que a tecnologia dos sistemas especialistas, através de algumas técnicas de IA, tenta imitar (**TROLLIP 91**).

Nos tempos atuais, as organizações têm no conhecimento acumulado de seu quadro de pessoal, um patrimônio muito importante para seus empreendimentos futuros e para sua própria sobrevivência face às constantes concorrências de mercado. Processos como “downsizing”, reengenharia e aposentadorias prematuras eram utilizados

constantemente pelas grandes instituições, comprometendo a memória institucional da empresa (LIEBOWITZ 95).

Ainda hoje, funcionários experientes aposentam-se ou deixam as organizações, resultando em perda incalculável pois o conhecimento de vários anos de experiência e treinamentos destes indivíduos serão perdidos, a menos que haja uma maneira de captar e preservar estas experiências dentro da empresa.

A tecnologia dos sistemas especialistas é ideal para atingir estes propósitos pois os processos e experiências funcionais da organização podem ser documentados através do uso de sistemas especialistas e a tecnologia do conhecimento poderá ser utilizada para capitalizar o conhecimento como um produto estratégico para as corporações.

As pesquisas em sistemas especialistas têm ainda um longo caminho a percorrer, particularmente avançando o estado da arte em aprendizagem de máquina, programação automatizada, e padronização. Mesmo assim, várias organizações já estão fazendo grandes economias e estão obtendo muitos benefícios na utilização desta tecnologia.

Portanto, existem ainda vários problemas relacionados ao desenvolvimento destes sistemas. Um deles é a incapacidade de um sistema em gerar um raciocínio pedagógico inteiramente autônomo, o que possibilitaria ao sistema tomar decisões que não tivessem sido antecipadas pelos especialistas, assim, os projetos são baseados em modelos que podem ou não representar o processo de aquisição de conhecimento do aluno.

No desenvolvimento do SESTATNET a ausência de um raciocínio pedagógico autônomo é compensada por um projeto pedagógico que, acompanhado pelos professores de estatística e uma pedagoga buscam conferir essas habilidades ao SESTATNET.

Como acontece exatamente aos sistemas especialistas em sua concepção a base de conhecimento é suprida inicialmente com o conhecimento dos especialistas do domínio.

O módulo **ACP** segue a concepção pedagógica do **SESTATNET** e foi desenvolvido utilizando as técnicas de Sistemas Especialistas.

4 – DESENVOLVIMENTO DO MÓDULO ACP

O desenvolvimento deste item terá por base NAKAZAWA e MARAFON 2003.

O software SESTATNET foi desenvolvido utilizando a plataforma Java 2 (POTTS-96), mais especificamente utilizando J2EE™.

A tecnologia J2EE™ tem seu modelo baseado em componentes, simplificando assim seu desenvolvimento. Controla a infra-estrutura e a interoperabilidade das aplicações, suportando serviços WEB com um desenvolvimento seguro e robusto.

Dentre os vários componentes que formam essa tecnologia, utilizamos apenas JSP (VIS-02), JavaBeans, JDBC e Java Servlets para o desenvolvimento do SESTATNET.

4.1 - Java Server Pages

Dentre as diversas tecnologias existentes que permitem a elaboração de *websites* com conteúdo dinâmico (PHP – *Personal Home Page*, ASP – *Active Server Pages*, Delphi, CGI-*Common Gateway Interface*), a equipe de desenvolvimento do SESTATNET optou pelo componente JSP (JSP-02) por se tratar de uma tecnologia para desenvolvimento de aplicações WEB que possui grande portabilidade de plataforma, podendo ser executado em diversos sistemas operacionais. Ela permite o desenvolvimento de aplicações que acessem banco de dados, arquivos-texto, capte informações a partir de formulários, informações sobre o visitante e sobre o servidor, a geração de gráficos dinâmicos, o uso de variáveis e *loops* entre outras coisas.

Além disso, essa tecnologia permite separar a programação lógica (parte dinâmica) da programação visual (parte estática), facilitando o desenvolvimento de aplicações mais robustas.

As páginas JSP (**JSP-03**) são páginas HTML (*Hypertext Markup Language*) com algumas *tags* embutidas, que podem conter código Java. A primeira vez que uma página JSP é processada, a JSP *engine* cria um código-fonte para um Java *Servlet* e então esse

código é compilado. Após essa compilação, dentro do servidor JSP existirá um *Java Servlet* que poderá ser acessado toda vez em que for requisitado.

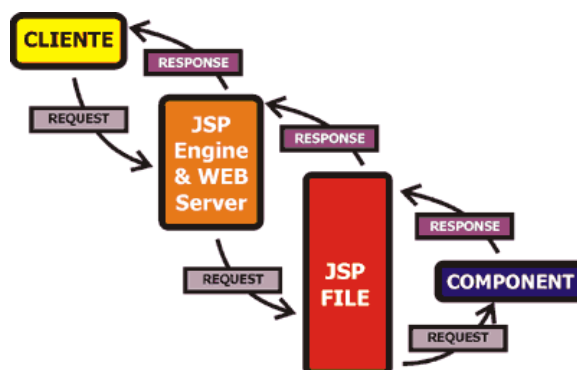


Figura 4 - Requisição de uma página JSP.

O processamento da requisição de uma página JSP funciona como a figura 4. O cliente faz a solicitação de um arquivo JSP, é enviado um *object request* para a *JSP engine*. A *JSP engine* envia a solicitação de um *JavaBean component* especificado no arquivo. O componente controla a requisição possibilitando a obtenção de um gráfico, geração de dados como resultado de uma análise feita ou acesso a um arquivo em banco de dados, em seguida, passa o objeto *response* de volta para a *JSP engine*. A *JSP engine* e o *WEB server* enviam a página JSP revisada de volta para o cliente, onde o usuário pode visualizar os resultados através do *WEB browser*.

Todas as páginas do SESTATNET “Estamos chamando de SESTATNET” que trocam informações com os alunos na análise estatística dos dados são JSP e realizam a comunicação direta com os *JavaBeans*, atuando como uma interface para os resultados estatísticos gerados por esses. Algumas páginas fazem uso de *JavaScript*³, que é responsável por encaminhar para uma determinada página dependendo da escolha do usuário.

A sequência da aprendizagem que o aluno segue se dá por meio de uma série de páginas, onde cada uma é *linkada* em uma ou mais páginas, formando assim os diversos caminhos de aprendizagem e que são armazenados no servidor.

4.2 - Classes do SESTATNET

Será dado foco maior à descrição das classes consideradas mais importantes no sistema (figura 5). As classes Administrador e Estatística serão descritas a seguir sendo elas as principais do SESTATNET.

Essas classes também são especiais para o módulo **ACP** pois serão nelas que os métodos criados nesta pesquisa serão implementados.

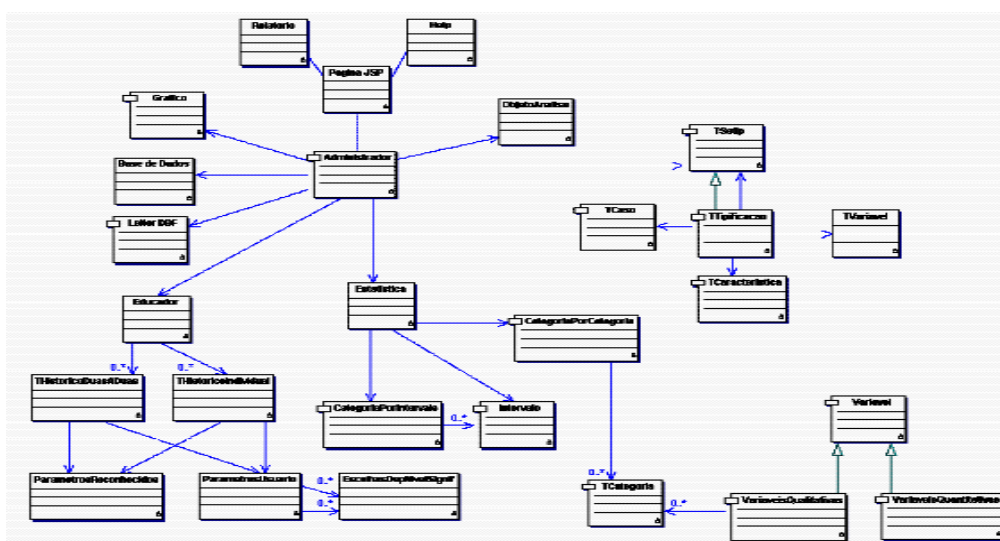


Figura 5 - Núcleo do Sistema SESTATNET

Para uma melhor visualização das classes e seus relacionamentos se faz necessário a explosão da figura 5 que descrevam melhor suas funções.

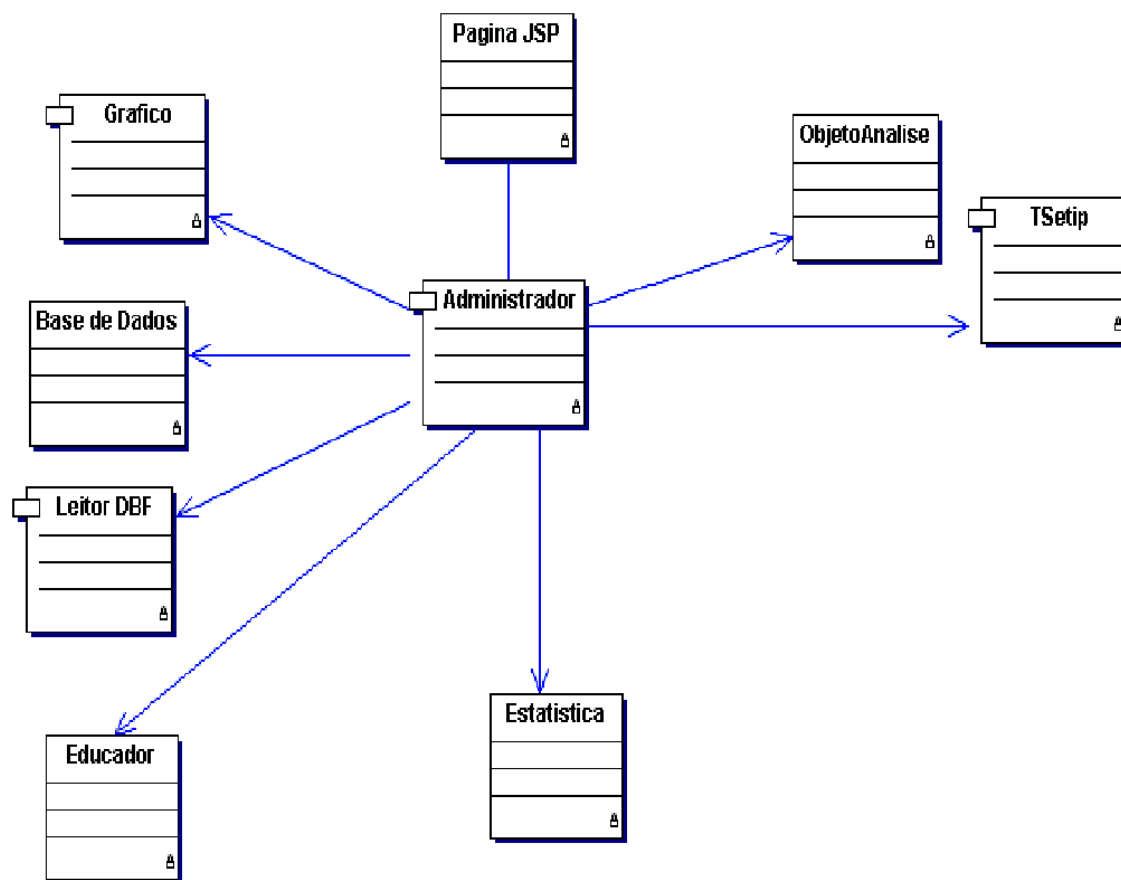


Figura 6 - Classe Administrador

4.2.1 – A Classe Administrador

Ponto central de todo o sistema é a classe responsável pela comunicação entre as páginas e as outras classes. Cada objeto desta classe é responsável pela gerência dos dados de um aluno específico, e por isso o seu *Bean* é declarado no escopo *session*, fazendo com que cada aluno tenha uma instância particular dessa classe para se relacionar com o sistema.

Todas as requisições das páginas JSP serão feitas diretamente para um *Bean* desta classe, o qual se encarregará de tratar tal requisição e realizar chamadas de funções das outras classes caso sejam necessárias.

Exemplo de uma chamada para esse *Bean*:

- Uma requisição da página JSP é enviada para um *Bean* administrador (Anexo I), informando, por exemplo, que o aluno escolheu a mensuração “Contínua” para uma descrição Univariada Quantitativa.
- O Administrador verifica se a variável selecionada é verdadeiramente “Contínua”, repassando uma mensagem para o aluno caso o sistema tenha reconhecido essa variável como “Discreta”. Importante verificar aqui, caso o sistema não concorde com a escolha do aluno, que o tipo de caminho escolhido pelo aluno fica em aberto, deixando para esse a oportunidade para continuar a sua análise ou voltar e escolher essa variável como “Discreta”.
- Verificado que o aluno escolheu a variável corretamente, ou que esse deseja continuar sua análise, será então executado o método “DescricaoUnivariadaQuantiCont” (Anexo II).

Método DescricaoUnivariadaQuantiCont:

- Primeiramente essa função requer alguns dados iniciais, como carregar do leitorDBF a variável Quantitativa escolhida pelo aluno, e informar essa variável para um *Bean* da classe Estatística para que esse possa analisá-la, informando assim as suas categorias.
- A partir dessas categorias, são retiradas informações que serão repassadas para um *Bean* gráfico.
- Também são gerados tabelas-resultados a partir de dados recolhidos do *Bean* Estatística, que futuramente irão ser passadas para a página JSP para ser exposto ao aluno.

Importante se verificar nessa seqüência de operações, que todas as informações passadas pelas páginas serão utilizadas apenas para processar os dados e armazená-los nos *Beans*, notando aqui que a classe Administrador age como um verdadeiro gerente do sistema.

Todas as informações requeridas pelas páginas JSP serão encaminhadas, também, somente para essa classe, como a requisição de uma tabela, ou dos dados para geração de um gráfico.

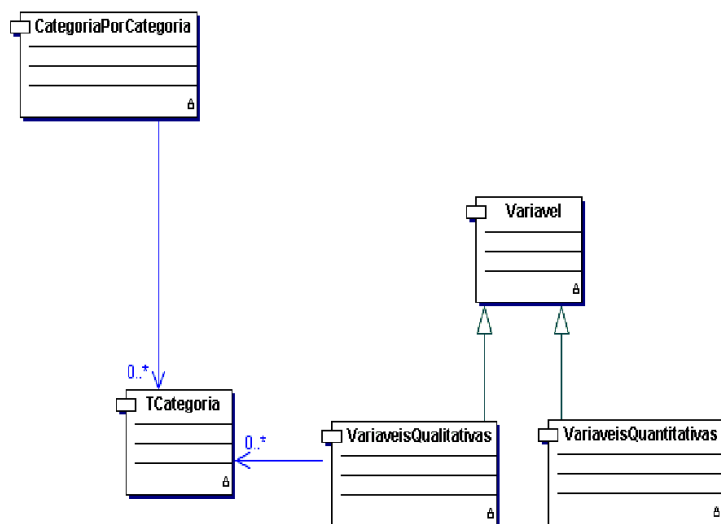


Figura 7 - Classe Variável

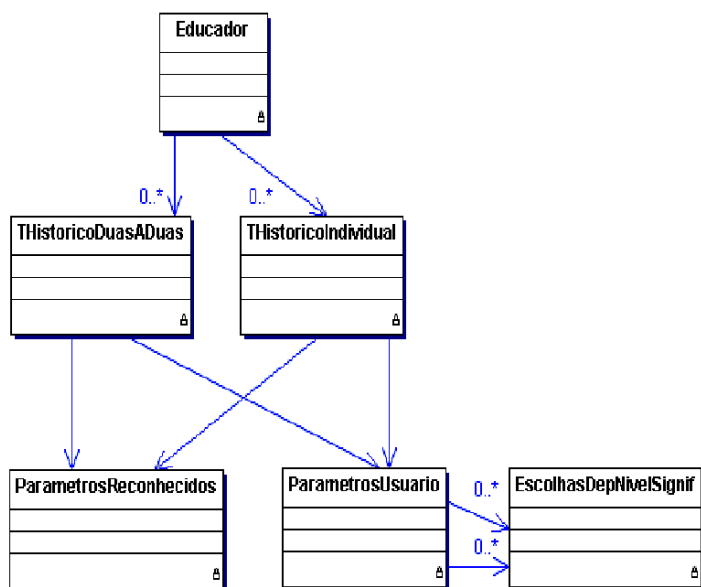


Figura 8 - Classe Educador

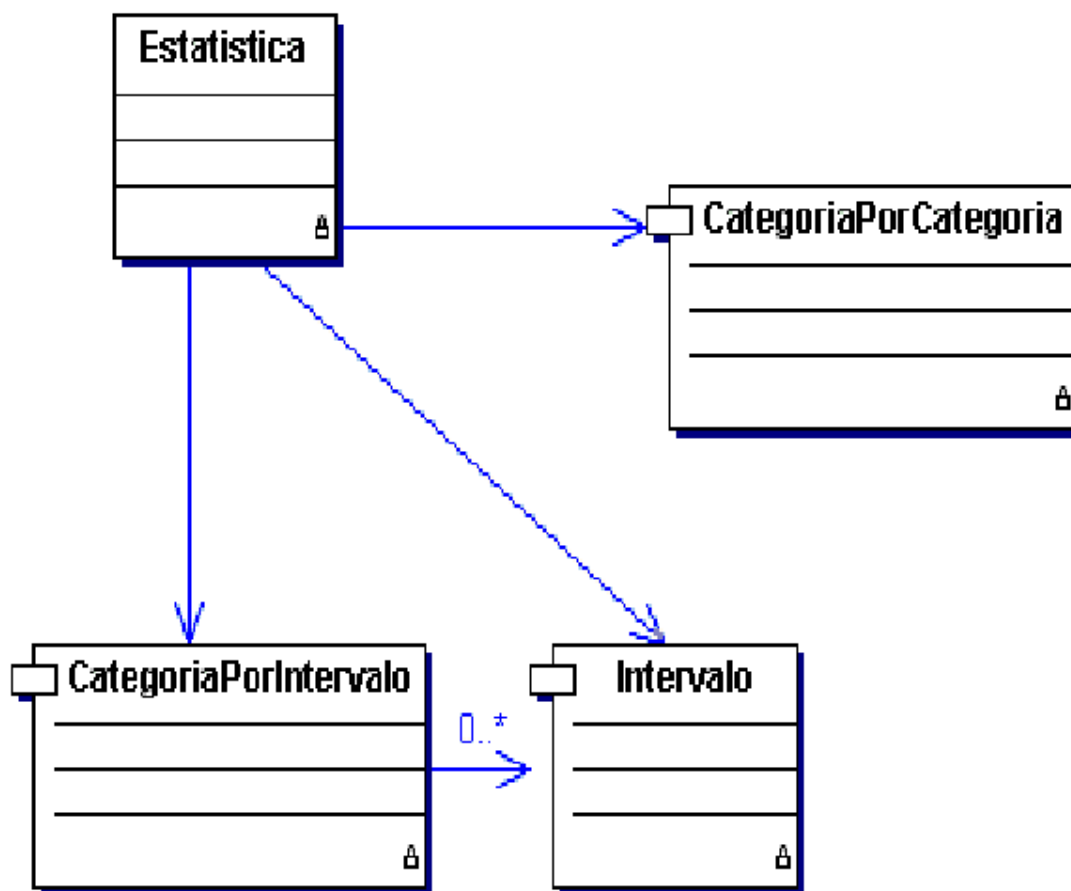


Figura 9 - Classe Estatística

4.2.2 – A Classe Estatística

Essa classe é responsável, principalmente pela geração e cálculo de todos os procedimentos estatísticos presentes no programa.

Os procedimentos estatísticos de inferência são: TesteT, TesteTSeparate, TesteTPares, Anova, QuiQuadrado, KruskalWallis, Wilcoxon, Spearman, Pearson, MannWhitney.

Os procedimentos de avaliação de suposições estatísticas são: frequência esperada, Homocedasticidade e Normalidade. Os procedimentos descritivos são: tabelas de frequências, gráficos e medidas descritivas.

Um objeto dessa classe é instanciado pela página JSP como sendo um *Bean* com escopo *application*, e encapsulado em seguida pelo *Bean* Administrador. Esse encapsulamento é realizado para se poder disponibilizar o uso de uma mesma instância para diversos alunos a fim de se economizar memória e processamento no servidor, ao mesmo tempo tornando a aplicação mais rápida pelo simples fato de não se necessitar a geração dessa nova instância.

É nessa classe que também serão implementados os métodos necessários aos cálculos exigidos pela análise por componentes principais, no desenvolvimento do módulo **ACP** nesta pesquisa.

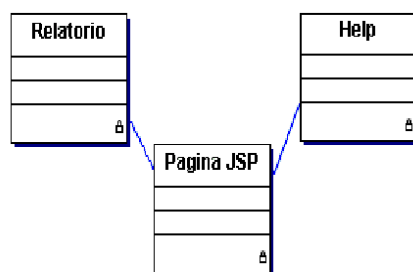


Figura 10 - Classe Relatório

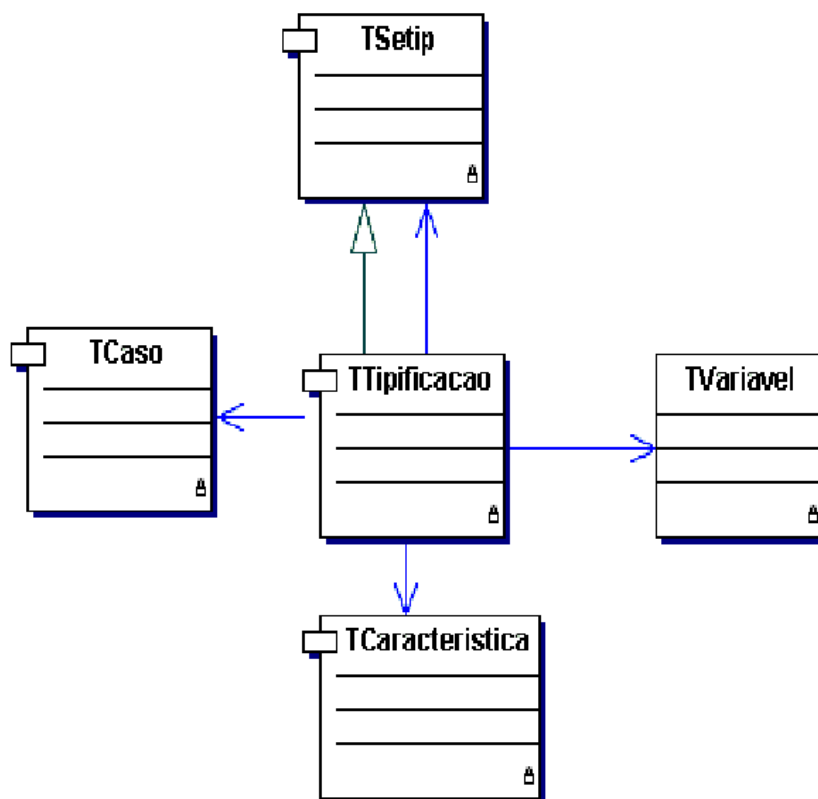


Figura 11 - Classe SETip

4.2.3 – A Classe SETip

A classe SETip é responsável pela tipificação das variáveis do sistema. Todos os módulos do SESTATNET utilizam-se dela, pois é o SETip quem identifica a variável trazendo os conceitos de variáveis quantitativas e qualitativas. Essa classe foi desenvolvida pelo aluno José Gonçalo em 2001, como dissertação de mestrado no INE-UFSC.

4.2.4 - Interface do SESTATNET

Nesse capítulo mostra-se a interface do SESTATNET. A sequência de figuras mostra um dos possíveis caminhos que o aluno pode seguir para a aprendizagem dos conceitos estatísticos.

A Figura 12 mostra a entrada do sistema, onde o aluno deve informar o seu *login* e sua respectiva senha.



The image shows the login interface of the SEstat.Net system. At the top, there is a blue banner with the text "SEstat.Net" and a Brazilian flag. Below the banner, the page is divided into two main sections. On the left, there is a sidebar with links: "Home", "help", "base", "docs", and "- Artigo". The "help" link is highlighted. On the right, the main content area is titled "Entrada no SEstat.Net". Below this title, there is a login form with the instruction "Digite seus dados para entrar". The form contains two input fields: "Usuário:" with the value "silvia" and "Senha:" with a masked password "XXXXXXXXXX". Below the input fields is a blue button labeled "Entrar". At the bottom of the page, there is a footer with logos for "Java" and "MySQL", and the text "2003 - LEA - Laboratório de Estatística Aplicada".

Figura 12 - Entrada do sistema. (NAKAZAWA e MARAFON 2003)

Após a entrada no sistema, são mostradas todas as bases de dados referentes ao aluno. Na figura 13 são mostradas as bases pap.dbf e teste.dbf que estão armazenadas no servidor e que podem ser utilizadas pelo aluno e este deve escolher uma delas para o seu trabalho.

SEstat.Net

Home
help
 - Criando uma base
 - Base PAP
 - Padrão dBase

base
docs
 - Artigo

Escolha da base de dados

Arquivo	Descrição	Criação	Última modificação
<input checked="" type="radio"/> pap.dbf	Base de Dados PAP	2003-01-16	2003-01-17
<input type="radio"/> teste.dbf	Base Teste	2003-01-17	2003-01-17

Avançar >

2003 - LEA - Laboratório de Estatística Aplicada

POWERED BY
 JAVA MySQL

Figura 13 - Escolha da base de dados do aluno. (AKAZAWA e MARAFON 2003)

Em seguida o aluno tem a opção de escolher qual o procedimento estatístico ele quer aplicar aos seus dados.

4.3 – O Desenvolvimento do módulo ACP

O SESTATNET é uma ferramenta de análise estatística de dados usando base de dados flexível, que disponibiliza aos alunos a realização de inferências estatísticas ou análises descritivas com uma ou duas variáveis, oferecendo os resultados estatísticos da análise, os possíveis caminhos de raciocínio e a linha de aprendizagem efetuada pelo usuário bem como a base de dados do aluno.

O software é composto por um “núcleo” que funciona como a parte inteligente do sistema, possuindo todas as funcionalidades necessárias para a administração dos dados e geração dos resultados. Esse “núcleo” é formado por classes Java, utilizando como interface para o usuário páginas JSP (FIELDS-00).

Por herdar a estrutura didático-pedagógica do SEstat, o SESTATNET também possui um sistema de *help* sensível ao contexto (MELO-03), onde conforme o ponto que o aluno se encontra são disponibilizados tópicos de ajuda que o auxiliam num melhor entendimento do contexto atual.

Os dados referentes aos alunos (login, senha, etc.) são armazenados num Banco de Dados *MySQL* no servidor.

Cada aluno possui uma conta no servidor onde ficam armazenadas todas as suas bases de dados. E essas são arquivos criados pelo próprio aluno no padrão *dBase* (.dbf), a partir de dados coletados de pesquisas.

O desenvolvimento do **ACP** foi realizado levando em consideração a modularização do **SESTATNET** e foi realizado nas seguintes fases:

- Estudo das classes do SESTATNET;
- Criação dos métodos necessários para obtenção dos resultados necessários a ACP.
- Implementação das classes e telas;

Após a realização dessa análise foi visto que não havia a necessidade da implementação de novas classes.

O próprio **SESTATNET** foi concebido com uma arquitetura que prevê constantemente o aumento na sua base de conhecimento, e a criação de novos módulos já era do conhecimento dos primeiros a implementarem o **SESTATNET**. (NAKAZAWA e MARAFON 2003).

As funções criadas para a implementação do módulo de ACP constam nos anexos dessa pesquisa.

A concepção pedagógica do **ACP busca** permitir ao aluno ser o construtor de sua aprendizagem de descobertas de padrões de associação entre variáveis quantitativas.

No SESTATNET / **ACP** :

- São oferecidos conceitos de análise por componentes principais;
- São oferecidos conceitos estatísticos necessários ao complemento das análises;
- São realizadas perguntas sobre a característica das variáveis;
- São disponibilizados resultados estatísticos (gráficos e numéricos);
- São utilizados algoritmos específicos para os cálculos da análise por componentes principais.

4.3.1 - Validação dos algoritmos.

Os algoritmos pesquisados relativos a análise por componentes principais tiveram grande utilidade nessa análise, mas os algoritmos avaliados do pesquisador John M. Lynch (**LYNCH 2003**) foram de grande valia. Após uma adequação e reorganização de procedimentos os algoritmos tiveram êxito em todos os testes. Este parágrafo está confuso!

Após adaptações paramétricas nos algoritmos os resultados foram conclusivos e validados pelo software Statistica v 5

4.3.2. INTERFACE DO ACP

Na figura 14 a tela principal do sistema é apresentada para que o aluno possa escolher o método de análise mais apropriada para o seu caso. O módulo **ACP** e **ACM** são acessíveis pela opção “**Análise Multivariada**”.



Figura 14 - Tela de Apresentação integrada ao Módulo de ACP

Na Figura 15 tem-se, de acordo com a base de dados carregada para o sistema, a apresentação das variáveis que fazem parte da amostra que se deseja analisar. Então o aluno deve selecionar as variáveis que serão eleitas para a futura análise.

The screenshot shows the SEstat.Net web application. The header features the site name 'SEstat.Net' in large blue letters and a Brazilian flag graphic on the right. A left sidebar contains navigation links: 'Home', 'help', 'base', 'docs', and '- Artigo'. The main content area is titled 'Escolha das variáveis' (Choose the variables) and asks the user 'Quais são as variáveis que você deseja trabalhar?' (Which variables do you want to work with?). It includes a hint: '(utilize as teclas Ctrl ou Shift para selecionar mais de uma variável)' (use the Ctrl or Shift keys to select more than one variable). A list of variables is displayed in a box: CASO, LOCAL, PAP, INSTR, TAM, RENDA88, RENDA98, ALIM98, APROV88, and APROV98. The 'TAM' variable is currently selected, highlighted in blue. Below the list is a button labeled 'Avançar >' (Advance >). The footer indicates the year 2003 and the institution LEA - Laboratório de Estatística Aplicada.

SEstat.Net

Home
help
base
docs
- Artigo

Escolha das variáveis

Quais são as variáveis que você deseja trabalhar?
(utilize as teclas Ctrl ou Shift para selecionar mais de uma variável)

- CASO
- LOCAL
- PAP
- INSTR
- TAM**
- REND88
- REND98
- ALIM98
- APROV88
- APROV98

Avançar >

2003 - LEA - Laboratório de Estatística Aplicada

Figura 15 - Seleção de Variáveis

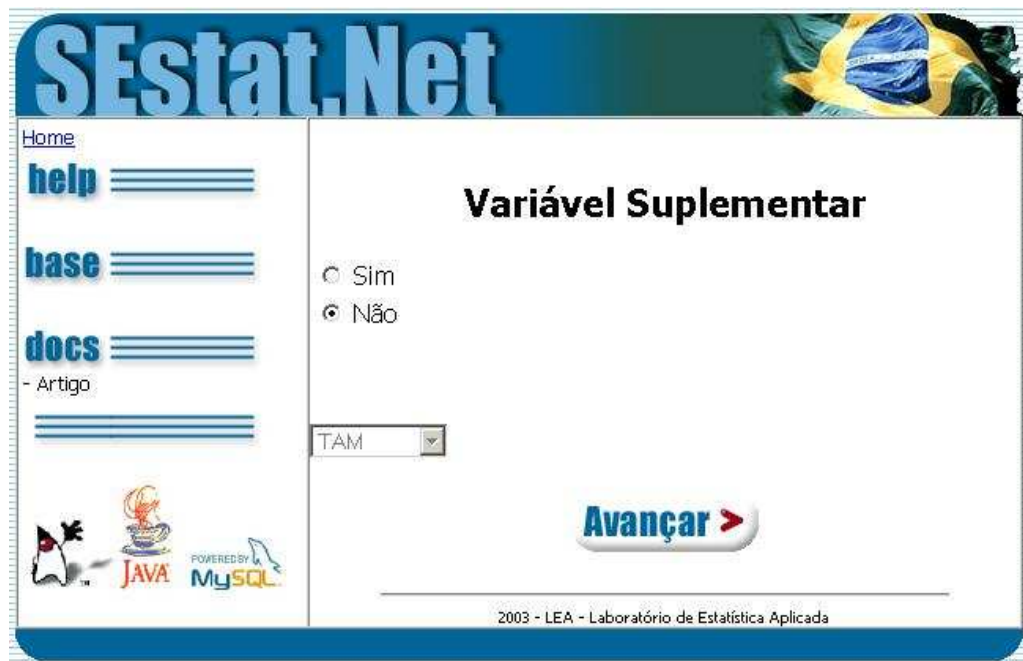
Na figura 16 o aluno deve fazer a qualificação das suas variáveis para análise. Já está presente nesse momento o SETip que servirá como um classificador para o sistema. Ele será utilizado para validarmos as respostas do aluno quanto a classificação das variáveis. Caso o aluno faça alguma classificação equivocada o SETip irá responder e orientar o aluno quanto a natureza daquela variável.

A classificação é necessária, pois através dela o SESTATNET irá fazer uso de um tipo de análise multivariada específica.

The screenshot shows the SEstat.Net web interface. The header features the site name 'SEstat.Net' and a Brazilian flag. The left sidebar contains navigation links: 'Home', 'help', 'base', 'docs', and '- Artigo'. The main content area is titled 'Tipo das variáveis' and prompts the user to 'Selecione o tipo da variável:'. It features two columns: 'Variáveis selecionadas:' on the left and classification options on the right. The 'Variáveis selecionadas:' column contains a list with 'RENDA98' and 'ALIM98', and a 'Reset' button below it. The classification options are organized into two sections: 'Qualitativa' and 'Quantitativa'. The 'Qualitativa' section has a 'Qualitativa' button. The 'Quantitativa' section has a 'Quantitativa' button and a list containing 'TAM' and 'RENDA88'. At the bottom of the main area is an 'Avançar >' button. The footer indicates '2003 - LEA - Laboratório de Estatística Aplicada'.

Figura 16 - Tipificação de Variáveis

A utilização da variável X complementar no módulo **ACP** terá grande utilidade quando os gráficos de sobreposição dos indivíduos em relação a variável complementar. Na figura 17 mostra como o aluno pode fazer a sua seleção.



The screenshot shows the SEstat.Net web application. The header features the 'SEstat.Net' logo and a Brazilian flag. On the left, there is a navigation menu with links for 'Home', 'help', 'base', 'docs', and 'Artigo'. The main content area is titled 'Variável Suplementar' and contains two radio buttons: 'Sim' (unselected) and 'Não' (selected). Below these is a dropdown menu currently set to 'TAM'. A large 'Avançar >' button is positioned at the bottom right of the form. The footer of the page indicates '2003 - LEA - Laboratório de Estatística Aplicada'.

Figura 17 - Seleção da Variável Suplementar

Como foi avisado ao aluno ele provavelmente fez uma seleção de variável X tipo de forma equivocada. Nessa hora o SESTATNET interage com o aluno por meio do **ACP** e o SETip integrado. Na Figura 18 mostra como é o aviso ao aluno.

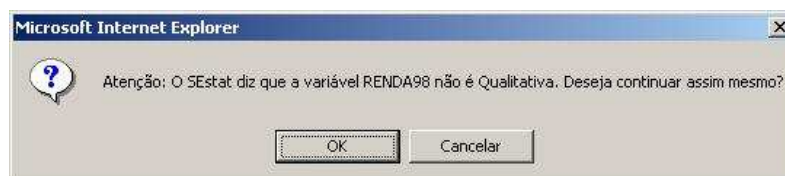


Figura 18 - Aviso do Sistema em caso de escolha errada

O primeiro resultado a ser avaliado na análise por componentes principais é o uma análise isolada do comportamento de cada variável. Medidas descritivas e gráficos são apresentados a fim de se oferecer a possibilidade de uma pré-análise. Um dos principais motivos de tal cuidado é a identificação visual e numérica de possíveis pontos discrepantes. Na figura 19

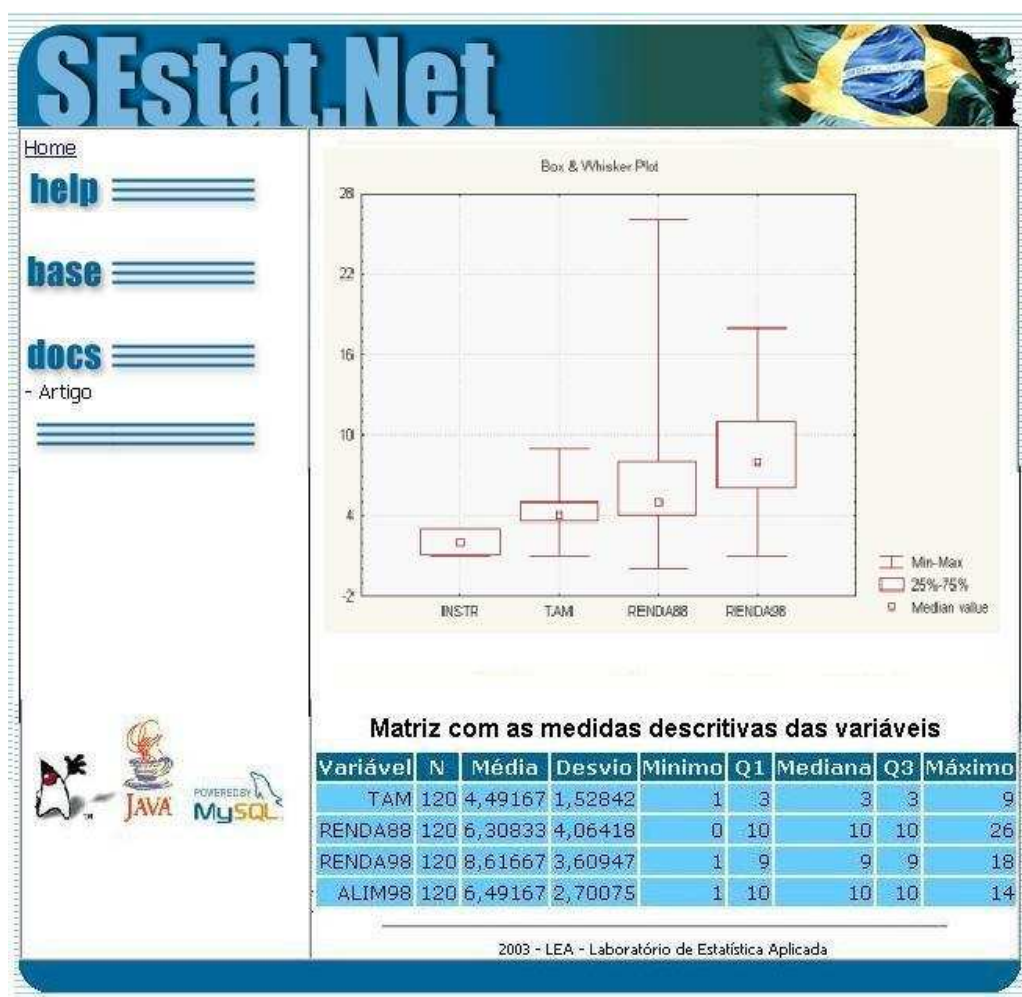


Figura 19 - Tela de apresentação de características descritivas

Na figura 20 é apresentada a matriz de correlação que foi gerada a partir dos valores lidos da base de dados do aluno. Essa matriz é muito importante para a análise por componentes principais, pois é ela que gera os autovalores e autovetores do sistema, isso sem avaliar toda a informação que a própria correlação oferece.

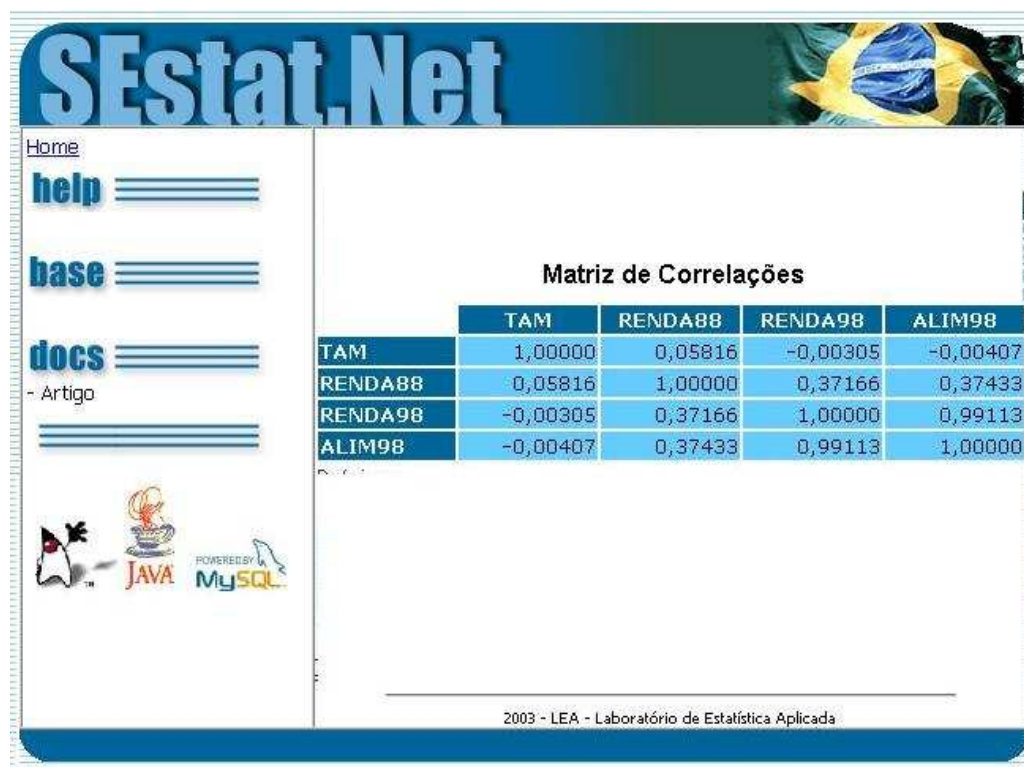


Figura 20 - Matriz de Correlação

Os autovalores são apresentados na figura 21, juntamente com a variância observada em cada um dos fatores.

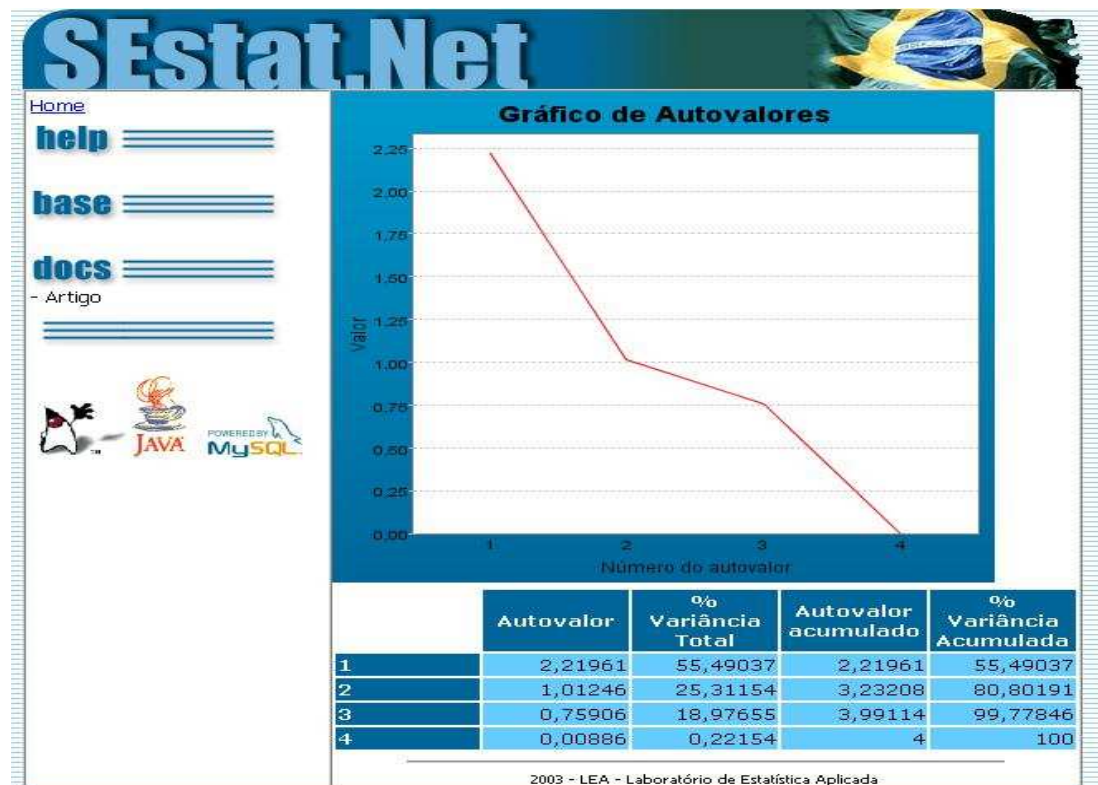


Figura 21 - Autovalores - Valores numéricos e gráfico de fatores

Na figura 22 são apresentadas às cargas fatoriais avaliadas de todos os fatores calculados.

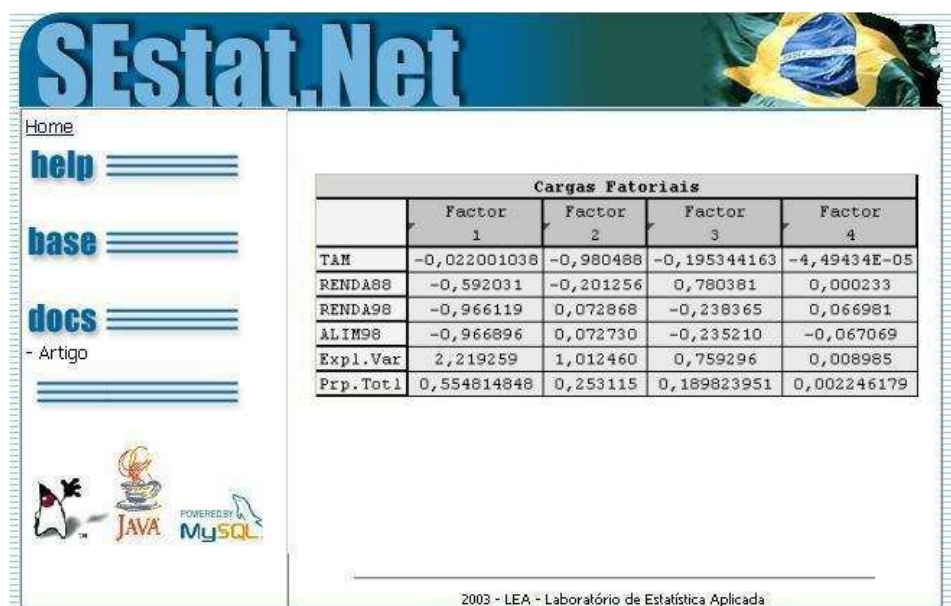


Figura 22 - Cargas Fatoriais

Na figura 23 são apresentados os coeficientes dos escores fatoriais. Esses são muito importantes pela geração posterior de matrizes que descreverão o comportamento gráfico das variáveis nos fatores e também dos indivíduos nos fatores.

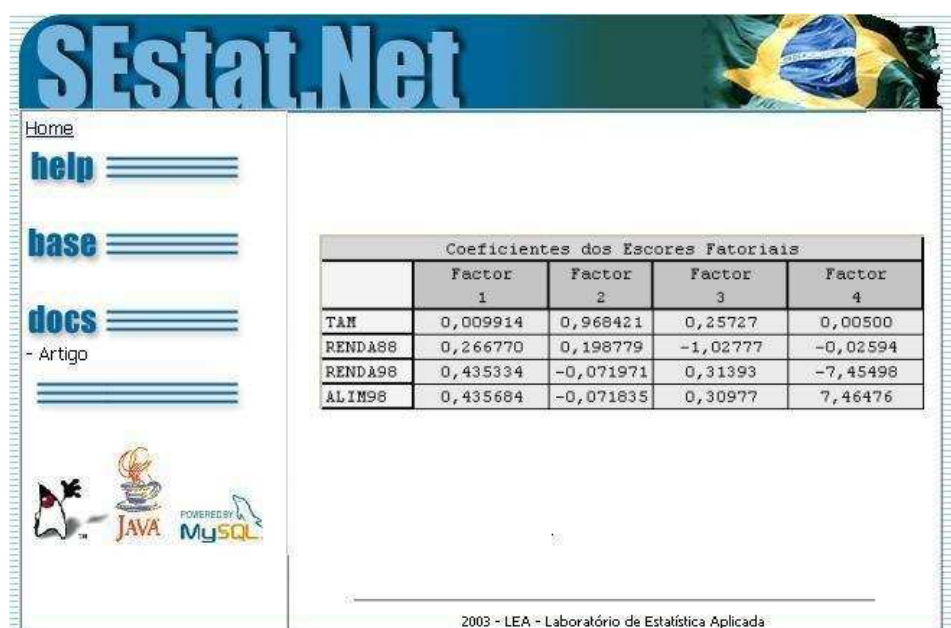


Figura 23 - Coeficientes de Escores Fatoriais

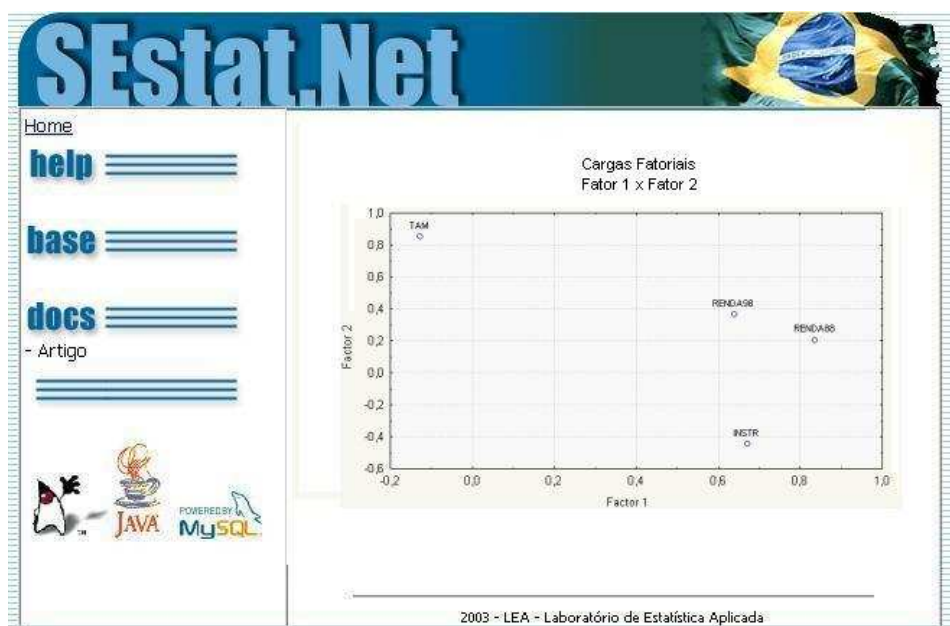


Figura 24 - Cargas Fatoriais - Fator 1 x Fator 2

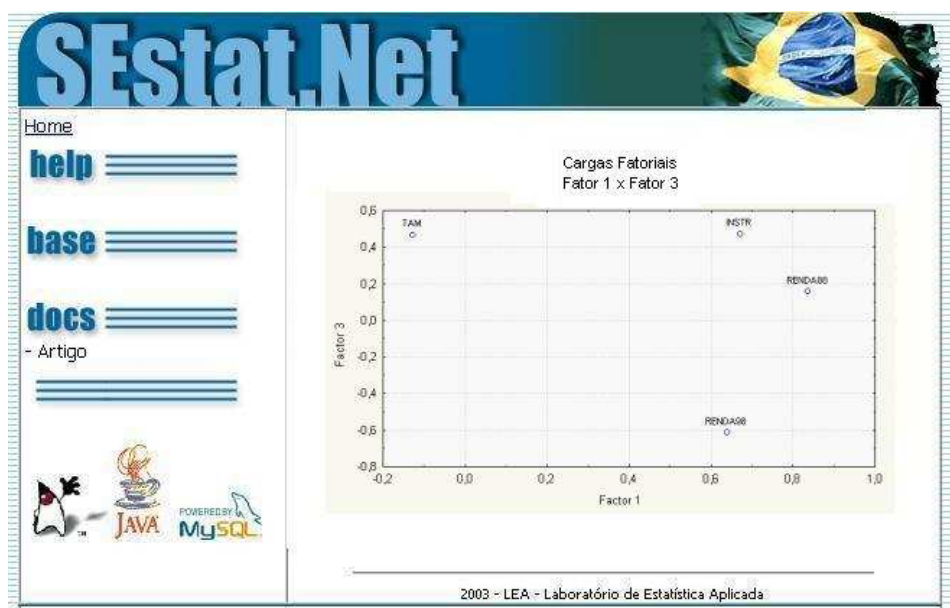


Figura 25 - Cargas Fatoriais - Fator 1 x Fator 3

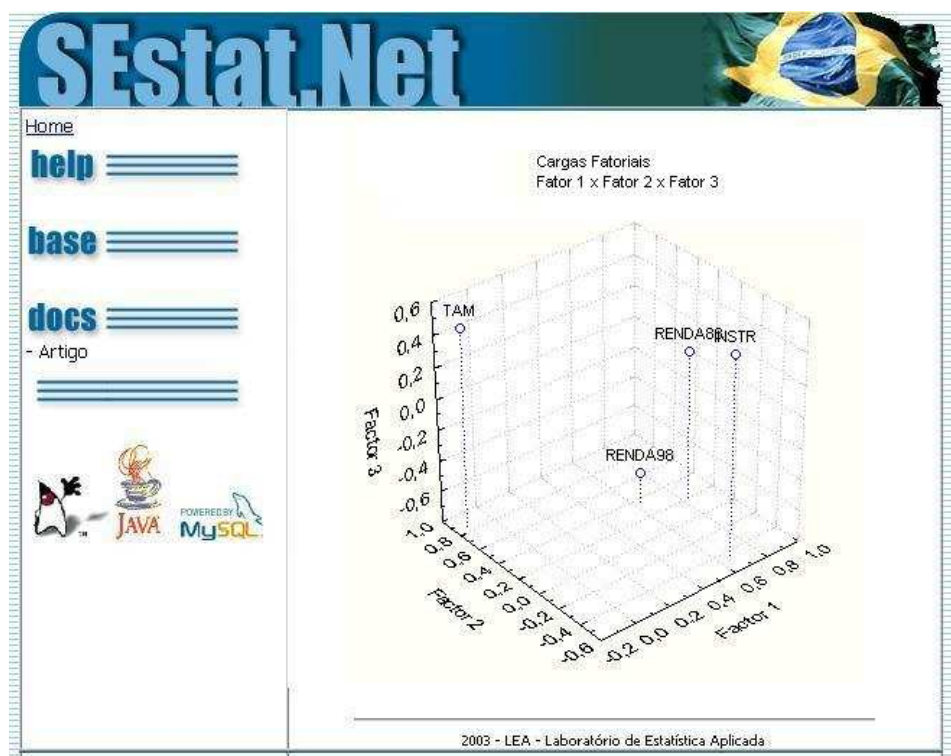


Figura 26 - Cargas Fatoriais - Fator 1 x Fator 2 x Fator 3

Nas figuras 24,25 e 26 foram mostradas representações gráficas das variáveis estudadas plotadas em gráficos 2D e 3D.

A matriz de escores fatoriais dos indivíduos é importante, pois aponta para o valor que cada um dos casos possui dentro do fator avaliado. Essa matriz servirá para plotar um gráfico que mostre todos os indivíduos ao mesmo tempo, bem como o destaque necessário para uma variável suplementar, completando os recursos oferecidos para o estudo do comportamento e características dessas variáveis.



The screenshot shows the SEstat.Net web application. On the left is a navigation menu with links for Home, help, base, docs, and - Artigo. Below the menu are logos for Java and MySQL. The main content area displays a table titled 'Escores Fatoriais dos Indivíduos' (Factorial Scores of Individuals). The table has 5 columns: Case, Factor 1, Factor 2, Factor 3, and Factor 4. It contains 21 rows of data, numbered 1 to 21. At the bottom of the page, it says '2003 - LEA - Laboratório de Estatística Aplicada'.

Case	Factor 1	Factor 2	Factor 3	Factor 4
1	0,85910	-0,92772	-0,43677	1,98835
2	-0,59903	0,00742	-1,61667	0,55781
3	-0,05002	0,31244	0,14048	-2,14705
4	1,25745	-0,47835	0,61294	-1,42784
5	-1,16288	0,10050	-2,02002	-0,83932
6	0,38034	1,11509	-0,58803	0,59465
7	-1,20915	-0,19059	-0,55392	-0,10884
8	1,25942	-0,00065	-1,78576	1,95333
9	2,32056	-0,57158	-2,38833	-0,80914
10	-0,24276	-0,35014	0,13807	-0,07855
11	0,46108	-0,70534	1,84160	-0,02272
12	0,09192	0,52802	-0,95803	-0,10718
13	1,41007	-0,14501	-1,07832	2,66465
14	1,81261	-0,21149	-0,78967	1,29782
15	0,65160	0,95079	0,20638	-0,75941
16	-0,17072	-1,81063	0,76175	0,63262
17	0,51375	-1,92364	1,25207	-0,03565
18	-0,74106	-2,35116	0,19008	-0,76778
19	0,06828	-0,85696	-0,70193	-2,16636
20	-0,26862	-0,74157	-1,24347	-0,80591
21	-0,57966	-0,23475	-0,40346	1,28190

Figura 27 - Escores Fatoriais dos Indivíduos

O ACP como um módulo de apoio a aprendizagem da análise por componentes principais será testado em cursos de estatística no CTC/UFSC a partir do semestre 2003.1 de onde poderão ser aprimoradas suas interfaces visando uma melhor produtividade na aprendizagem, segundo os conceitos e pontos pedagógicos que poderão ser destacados.

5. – Conclusão.

Este trabalho abordou a utilização de sistemas especialistas como uma forma viável de desenvolvimento de sistemas educacionais. Utilizando essa técnica de inteligência artificial foi criado um módulo novo para o SESTATNET aumentando sua base de conhecimento.

A utilização de meios computacionais comprovou sua eficácia no ensino e juntamente com os devidos aspectos pedagógicos tornam-se agentes facilitadores e ampliadores no processo de aprendizagem.

Foi desenvolvido o protótipo de um módulo chamado ACP – Módulo de análise por componentes principais que integra o SESTATNET (Sistema especialista de ensino de estatística).

Na sua implementação foram utilizados a plataforma Java 2 da Sun Microsystem. As bases de dados que podem ser acessadas pelo sistema são do tipo DBF – Data base File e também é utilizado o servidor de banco de dados Mysql como apoio ao processo de análise.

Como resultados observados podem-se destacar os seguintes:

- Através de simulações no sistema foram verificadas que os cálculos efetuados pelos algoritmos implementados tiveram êxito, pois foram verificados por rotinas semelhantes no software Statistica.
- O problema da complexidade da análise multivariada foi solucionado de forma mais natural do que as soluções convencionais (pacotes estatísticos).

Este trabalho apresenta muitas possibilidades de continuação como, por exemplo:

- A possibilidade de seleção e visualização de mais de uma variável complementar dentro da análise.
- A implementação de novos métodos de extração.
- Criação de novos gráficos que incluam a variável suplementar juntamente com as variáveis estudadas.

6 – Referências Bibliográficas

- AKHRAS**, F. N., SELF, J., “A Process-oriented Perspective on Analysing Learner-Environment Interactions in Constructivist Learning”, em Proceedings of the Brazilian Symposium on Computing in Education (SBIE’95), Florianópolis, 1995.
- ALMEIDA**, Prof. Prof. Tabajara Lucas de – Análise Fatorial –, Departamento de Matemática – FURG, disponível em <http://lula.dmat.furg.br/~taba/posanafat.htm>. Acesso em 02 de Março de 2003.
- SODRÉ**, Ulysses **2004** – Álgebra Linear – Autovalores e Autovetores – disponível em <http://pessoal.sercomtel.com.br/matematica/superior/álgebra/autoval/autoval.htm>. Acesso em 23 de Novembro de 2004.
- ALVES**, Rubem. *Filosofia da Ciência*. São Paulo, Brasiliense, 1981.
- BANET**, Tomás Aluja e **MORINEAU**, Alain. *Aprender de los datos: el análisis de componentes principales*. Barcelona, Ediciones Universitarias de Barcelona,, 1999.
- BUCHANAN**, B.G., **FEIGENBAUM**, E.A., “DENDRAL and Meta-DENDRAL: Their Applications Dimension” *Artificial Intelligence*, 11(1,2), pp. 5-24, 1978.
- BUCHANAN**, **BUCHANAN**, B.G., **SMITH**, R.G., “Fundamentals of Expert Systems” em **BARR**, A., **COHEN**, P.R., **FEIGENBAUM**, E.A., “The Handbook of Artificial Intelligence Vol. IV” Addison-Wesley Publishing Company, Inc., 1989.
- CARVALHO**, Maria Alice Pessanha - Análise de um Ambiente Construtivista de Aprendizagem a Distância: Estudo da Interatividade, da Cooperação e da Autonomia no Curso de Gestão Descentralizada de Recursos Humanos em Saúde / Maria Alice Pessanha de Carvalho. Rio de Janeiro: UFRJ/NUTES, 2000.
- CHAIBEN**, Hamilton – Inteligência Artificial na Educação, Disponível em <http://www.cce.ufpr.br/~hamilton/iaed/iaed.htm>. Acesso em 05 de Março de 2003.
- CHORAFAS**, Dimitris N – Sistemas Especialistas: aplicações comerciais ; tradução Mirian Fonseca Diniz; revisão técnica Nizam Omar. – São Paulo: McGraw-Hill, 1988.
- CIRIGLIANO**, G. F. L. La education abierta. Buenos Aires : El Ateneo, 1983.
- CLANCEY**, W.J., “Knowledge-Based Tutoring: The GUIDON Program”, The MIT

Press, 1987.

CLANCEY, W.J., “Methodology for Building na Intelligent Tutoring System”, em KEARSLEY, G., “Artificial Intelligence and Instruction - Applications and Methods”, pp. 193-227 , 1987.

COHEN, P.R. e **FEIGENBAUM**, E.A., “The Handbook of Artificial Intelligence - Vol. III”, Addison-Wesley Publishing Company, Inc., 1989.

ESCOFIER, Brigitte e **PAGÈS**, Jérôme..*Análisis factoriales simples y múltiples*. Paris, Dunod, 1990.

FOWLER, D.G., “A Model for Designing Intelligent Tutoring Systems”, Journal of Medical Systems, Vol. 15, N.1, 1991.

GARCIA, Othon M. *Comunicação em prosa moderna*. 17 ed.Rio de Janeiro, Editora da Fundação Getúlio Vargas, 1996.

GARRISON, D. R. Three generations of techonological innovation in distance education. Distance Education, n. 6, 1985.

GUÉDEZ, V. Las Perspectivas de la educación a distancia en el contexto de la Educación Abierta y Permanente. Boletín Informativo da la Asociación Iberoamericana de Educación Superior a Distancia, n. 3, Madrid, 1984.

HAIR, Joseph F. Multivariate Data Analysis : with readings. 4.ed. New Jersey : Prentice-Hall, 1995.

HOLMBERG, B. Distance education a survey and bibliography. London : Kogan Page, 1977.

JOHNSON, Richard Arnold. Applied multivariate statistical analysis Richard A. Johnson. Dean W. Wichers. – 4th ed. Prentice-Hall: EUA, 1998

JONASSEN, D. O uso das novas tecnologias na educação a distância e aprendizagem construtivista. Em Aberto, v. 16, n. 70, p.70-88, 1996.

JONASSEN, D.H., **WANG**, S., “The Physics Tutor: Integrating Hypertext and Expert Systems”, Journal of Educational Technology Systems, Vol. 22(1), pp. 19-28, 1993.

JUN, Yu - Projeto de iniciação científica - Autovalores e autovetores; Teorias e aplicações. Instituto de Matemática, Estatística e Computação Científica (IMECC) Unicamp. Orientadora: Márcia Aparecida Gomes Ruggiero, disponível em [http : // sites . uol.com.br/matmatik/Iniciacao_Cientifica/projeto.html](http://sites.uol.com.br/matmatik/Iniciacao_Cientifica/projeto.html), Acesso em 07 de Março de 2003.

- KEMP, R.**, “Intelligent Computer Assisted Instruction: A Knowledge-Based Perspective”, The Australian Computer Journal, Vol. 24 N. 3, pp.121-129, 1992.
- KIRLIAM 2000**, Maciel Dias - trabalho de conclusão de curso : SEstat - Sistema Especialista para Auxílio ao Ensino de Estatística - Departamento de Informática e Estatística, Curso de Ciências da Computação - Florianópolis, dezembro de 2000.
- KNOWLES, M. S.** The modern practice of adult education from pedagogy to andragogy. Nova Iorque : Cambridge, 1980. 400p.
- LIEBOWITZ, J.**, ”Expert Systems: Dead or Alive ?”, Educational Technology, pp. 53-55, March-April 1995.
- LINARD, M., ZEILIGER, R.**, “Designing a Navigational Support for an Educational Software”, versão eletrônica em “<http://www.irpeacs.fr/papers/rz/artmosc.htm>”, 1995.
- LLAMAS, GARCIA, J. L.** Un modelo de analisis para la evaluación del rendimiento academico en la ensenanza a distancia. Madrid : OEI, 1986.
- LONGO**, Maurício B. e **SMITH JR**, Ronaldo. *Delphi 3 Total*. Brasport, 1999.
- LÓPEZ, OSCAR C.** - Sistemas Especialistas, Disponível em http://www.eps.ufsc.br/~oscar/exp_sys/ing/se_Intro_parte1.htm. Acesso em 03 de Março de 2003.
- LYNCH, John M**, disponível em <http://www.public.asu.edu/~jmlynch>, visitado em 01 de Março de 2003.
- MANLY, Bryan F. J.** *Multivariate statistical Methods*. Londres, Chapman & Hall, 1994.
- MCKENZI, EMCKENZIE, N.**; Postgate, B.; Scuphan, J. Enzerianza abierta. Sistemas de ensenanza postsecundaria a distancia. Madrid : UNESCO, 1979
- MINSKY, M.** (editor), “Semantic Information Processing”, The MIT Press, Cambrige, 1968.
- MINSKY, M.** , “The Society of Mind”, Simon & Schuster, Inc, New York, 86.
- MONTEIRO, Conceição P. e OLIVEIRA, Maria Helena C. de.** *Metodologia da linguagem*. São Paulo, Saraiva, 1985.
- MOORE, M.G.** Learner autonomy : the second dimension of independent learnig.

Virginia : Warrenton, 1972. p.212 (Collection of Conference Papers; v.2).

MULLER, Mary Stela e **CORNELSEN**, Julce Mary. *Normas e padrões para teses, dissertações e monografias*. 3 ed. Londrina, Editora da Universidade Estadual de Londrina, 2001.

NAKAZAWA e MARAFON, Carlos Alberto e **MARAFON**, Márcio Juliandrei. **SESTATNET - ENSINO DE ESTATÍSTICA MEDIADO POR COMPUTADOR**, Universidade Federal de Santa Catarina , Fevereiro de 2003.

NIEDERAUER, Juliano. Disponível em [http:// www . somatematica . com.br](http://www.somatematica.com.br). Acesso em 02 de Março de 2003.

NOAL, Renato Bica – **AMBIENTE COLABORATIVO PARA ENSINO DE ESTATÍSTICA COM O SESTAT** - Departamento de Informática e Estatística, Curso de Pós-graduação Ciências da Computação - Florianópolis, fevereiro de 2002.

PARK, O., **PEREZ**, R.S. and **SEIDEL**, F.J., “Intelligent CAI: Old Wine in New Bottles or a New Vintage ?”, em **KEARSLEY**, G., “Artificial Intelligence and Instruction – Applications and Methods” ;, pp. 11-45.

PEPER, G., “Hypertext: Its Relationship to, and Potencial Impact on, Knowledge Based Systems”, 1991.

PEREIRA, Julio César R. *Análise de Dados Qualitativos: Estratégias Metodológicas para as Ciências da Saúde, Humanas e Sociais* / Júlio César R. Pereira – São Paulo: Editora da Universidade de São Paulo, 1999.

PETERS, O. *Die didaktische strukur des fernunterrichts untersuchungen zu einner industrialisierten form des lehrens and lernens* . Weinhein : Beltz, 1973.

PILETTI, Nelson. *História da educação no Brasil*. São Paulo, Ática, 1994.

PIVETA, Jaqueline – **Módulo Inteligente de Análise de Correspondência Múltipla para o SESTATNET Trabalho Individual** apresentado em fevereiro de 2003. INE/USFC.

RICH , E., “Inteligência Artificial”, tradução do original “Artificial Intelligence”, McGraw-Hill, 1988.

RICHMOND, W.K., “A Revolução no Ensino” tradução do original “The Teaching Revolution”, Companhia Editora Nacional, 1975.

RICKEL, J.W., “Intelligent Computer-Aided Instruction: A Survey Organized Around System Components”, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 19, N. 1, pp. 40-57.

RUSSEL, Stuart J. *Artificial intelligence*. Rio de Janeiro, Prentice-Hall do Brasil, 1995.

SALGADO, Luiz Antonio. “Autovalores e Autovetores Associados de Matriz”. Disponível em http://www.lasalgado.eng.br/trabtec/sc2/apostila/cap7_6.htm, Acesso em 08 de Março de 2003.

SAVORY, S. E.(editor), ”Some Views on the State of Art in Artificial Intelligence” em “Artificial Intelligence and Expert Systems”, Ellis Horwood Limited, 1988, pp. 21-34, Inglaterra .

SCHUTZER , D. , “Artificial Intelligence - An Applications-Oriented Approach”, Van Nostrand Reinhold Company, New York, 1987.

SHORTLIFFE, E.H., “Computer-Based Medical Consultations: MYCIN”, New York: American Elsevier, 1976.

TINSLEY, Howard E.A. e **BROWN**, Steven D. *Handbook of applied multivariate statistics and mathematical modeling*. Flórida, Academic Press, 2000.

TROLLIP, S.R. & **ALESSI**, S.M., “Computer - Based Instruction , Methods and Development” Second Edition, Prentice Hall, INC, 1991.

VERDINELLI, Miguel Angel - Artigo - Análise Multivariada - Análise de Inércia – Florianópolis, setembro de 2000.

SCREMIN, Marcos Antônio Antonello - MÉTODO PARA A SELEÇÃO DO NÚMERO DE COMPONENTES PRINCIPAIS COM BASE NA LÓGICA DIFUSA - UFSC - Engenharia de Produção - Florianópolis - Março de 2003.

WIKIPEDIA – A enciclopédia Livre, disponível em <http://pt.wikipedia.org/wiki/Ead> , consultada em Agosto de 2006.

ANEXOS

Algoritmos

Abaixo seguem algoritmos em linguagem estruturada (Portugol) que serviram de base para a implementação dos métodos que foram adicionadas as classes do SESTATNET.

Os métodos descritos e adicionados são responsáveis pelos cálculos de análise multivariada, uma vez que para os cálculos estatísticos descritivos básicos o SESTATNET já possui métodos e classes apropriadas para tais tarefas. Seguem então os algoritmos para a determinação de Autovalores, Autovetores, Scores Fatoriais e a Matriz de Correlação.

É apropriado uma identificação de alguns elementos dos algoritmos utilizados:

R = Número de casos a serem analisados nas variáveis ativas selecionadas

C = Número de variáveis ativas na análise

S2 = Matriz de Autovalores

r2 = Matriz de Correlação

D8 = matriz com o percentual total da variância

D6 = matriz com o percentual acumulado de variância no fator

Após os procedimentos de calculo dos autovalores e autovetores e da Matriz de Correlação:

r2 = conterá a matriz com os Autovetores

b2 = conterá a matriz com as Cargas Fatoriais

v1 = poderá ser obtida o percentual de variância total por fator analisado utilizando a fórmula: $R2[J, I]^2 * D8[I]$

J = Indica o Componente

x2 = o Score do componente

Algoritmo em Portugol para a determinação de AutoValores e AutoVetores

```

inicio
  N <- R
  L <- C
  A <- 0.00000001
  C <- 0
  para I de 2 ate N faça
    para J de 1 até I - 1 faça
      C <- C + 2 * (R2[I, J] ^ 2)

```

```

    fim para
  fim para
  Y <- sqrt(C)
  O <- (A / N) * Y
  T <- Y
  D <- 0
  inicio
    T <- T / N
    para Q de 2 até N faça
      para P de 1 até Q - 1 faça
        se ABS(R2[P, Q]) >= T
          inicio
            D <- 1
            V <- R2[P, P]
            Z <- R2[P, Q]
            E <- R2[Q, Q]
            F <- .5 * (V - E)
            se F = 0
              g <- -1
            senao
              G <- -(SGN(F))
          fim se
          G <- G * Z / (sqrt(Z ^ 2 + F ^ 2))
          se G < -1 entao
            G <- -.999999
          fim se
          H <- G / (sqrt(2 * (1 + sqrt(1 - G ^ 2))))
          K <- sqrt(1 - H ^ 2)
          para I de 1 até N faça
            se (I = P) ou (I = Q) entao
              C <- B2[I, P]
              F <- B2[I, Q]
              B2[I, Q] <- C * H + F * K
              B2[I, P] <- C * K - F * H
            senao
              C <- R2[I, P]
              F <- R2[I, Q]
              R2[Q, I] <- C * H + F * K -> AUTOVALORES
              R2[I, Q] <- R2[Q, I] -> AUTOVALORES
              R2[P, I] <- C * K - F * H -> AUTOVALORES
              R2[I, P] <- R2[P, I] -> AUTOVALORES
            fim se
          fim para
          R2[P, P] <- V * K ^ 2 + E * H ^ 2 - 2 * Z * H * K -> AUTOVALORES
          R2[Q, Q] <- V * H ^ 2 + E * K ^ 2 + 2 * Z * H * K -> AUTOVALORES
          R2[P, Q] <- (V - E) * H * K + Z * (K ^ 2 - H ^ 2) -> AUTOVALORES
          R2[Q, P] <- R2[P, Q] -> AUTOVALORES
        fim
      fim para
    fim para
    se T > O
      d <- 0
    fim se
    D <- 0
  fim
  para I de 1 até N
    Q1[I] <- I
  fim para
  J <- 0
  inicio
    V1 <- 0
    J <- J + 1
    para I de 1 até N - J
      se R2[I, I] < R2[I + 1, I + 1] entao
        V1 <- 1
        V2 <- R2[I, I]
        R2[I, I] <- R2[I + 1, I + 1] -> AUTOVALORES

```

```

        R2[I + 1, I + 1] <- V2      -> AUTOVALORES
        P <- Q1[I]
        Q1[I] <- Q1[I + 1]
        Q1[I+1] <- P
    fim se
fim para
se V1 = 0
    fim
fim se
fim

Q <- 0
D7 <- 0
para J de 1 até N faça
    S2[J] <- R2[J, J]      -> AUTOVALORES
    se S2[J] > 0 entao
        S3[J] <- sqrt(S2[J])
        D7 <- D7 + (S2[J])
    senao
        S2[J] <- 0.00000001  -> AUTOVALORES
    end
fim para
para J de 1 até N faça
    D8[J] <- 100 * S2[J] / D7
    D9 <- D9 + (S2[J]) / D7
    D6[J] <- 100 * D9
    K <- Q1[J]
    V <- 0
    para I de 1 até N faça
        V <- V + B2[I, K] ^ 2
    fim para
    para I de 1 até N faça
        R2[I, J] <- B2[I, K] * sqrt(1 / V)  -> AUTOVALORES
    fim para
fim para
para I de 1 até N faça
    R2[I, 2] <- R2[I, 2] * (-1)  -> AUTOVALORES
    R2[I, 3] <- R2[I, 3] * (-1)  -> AUTOVALORES
fim para
para I de 1 até N faça
    para J de 1 até N faça
        C1[I, J] <- 0
        B2[I, J] <- 0
    fim para
    C1[I, I] <- S3[I]
fim para
para I de 1 até N faça
    para J de 1 até N faça
        para K de 1 até N faça
            B2[I, J] <- B2[I, J] + (R2[I, K] * C1[K, J])
        fim para
    fim para
fim para

```

Algoritmo em Portugol para determinação dos Escores fatoriais

```

C <- L
para I de 1 até C faça
    para J de 1 até R faça

```

```

        X1[I, J] <- 0
    fim para
fim para
para I de 1 até C faça
    para J de 1 até R faça
        para K de 1 até R faça
            X1[I, J] <- X1[I, J] + X2[K, I] * R2[K, J]
        fim para
    fim para
fim para

```

Algoritmo em Portugol para determinação da matriz de Correlação

```

início
para I de 1 até (r-1) faça
    para j de 2 até r faça
        x5 <- 0
        X6 <- 0
        Y5 <- 0
        Y6 <- 0
        Z5 <- 0
        para K de 1 até C faça
            X5 <- X5 + X2[I, K]
            X6 <- X6 + X2[I, K] ^ 2
            Y5 <- Y5 + X2[J, K]
            Y6 <- Y6 + X2[J, K] ^ 2
            Z5 <- Z5 + X2[I, K] * X2[J, K]
        fim para
        X4 <- (X6 - (X5 ^ 2) / C)
        Y4 <- (Y6 - (Y5 ^ 2) / C)
        Z4 <- (Z5 - (X5 * Y5) / C)
        R2[I, J] <- Z4 / (sqrt(X4 * Y4)) -> AUTOVALORES
        R2[J, I] <- R2[I, J] -> AUTOVALORES
    fim para
    R2[I, I] <- 1 -> AUTOVALORES
fim para
R2[R, R] <- 1 -> AUTOVALORES
para I de 1 até R faça
    A <- 0
    B <- 0
    para J de 1 até C faça
        Q <- X2[I, J]
        A <- A + Q
    fim para

```



```

        B <- B + Q ^ 2
    fim para
    B <- sqrt(ABS((B - A ^ 2 / C) / (C - 1)))
    A <- A / C
    para J de 1 até C faça
        X2[I, J] <- (X2[I, J] - A) / B
    fim para
fim para
para I de 1 até R faça
    para J de 1 até C faça
        X1[J, I] <- X2[I, J]
    fim para
fim para
para I de 1 até R faça
    R1[I] <- R2[I, I]
fim para
fim.

```

Métodos comuns ao SESTATNET

```

//*****
// Retorna o somatório de um número.
// Parâmetros:
//    __valor: valor que se deseja realizar o somatório.
//*****
public long getSomatorio(int __valor) {
    if (__valor <= 0)
        return 0;
    else
        return __valor + getSomatorio(__valor-1);
}

//*****
// Retorna o sinal de um determinado valor.
// Parâmetros:
//    __valor: valor que se deseja analisar.
//*****
protected int getSign(double _valor) {
    if (_valor < 0)
        return -1;
    else if (_valor > 0)
        return 1;
    else
        return 0;
}
//*****
// Acha o valor mínimo da variável, além de calcular a média desta variável.

```

```

// Parâmetros:
//    __valores: valores da variável.
//    __n: retorna o número de casos da variável.
//    __media: retorna a média da variável.
//    __mínimo: retorna o valor mínimo da variável.
//*****
public int AcheMenorValor(String[] __valores, Objeto __n, Objeto __media, Objeto __minimo) {
    long i,tamanho;
    double aux2,soma;
    int ret;
    tamanho = __valores.length;
    Objeto aux = new Objeto(0);
    Objeto j = new Objeto(0);
    ret = PrimeiroValor(__valores,aux,j);
    if(ret != 0){
        return(ret);
    }else{
        // aux contém o primeiro (não nulo) valor do array
        __n.setLong(1);
        soma = aux.doubleValue();
        for(i=j.intValue()+1; i<=tamanho-1;i++){
            if(!__valores[(int)i].equals("")){
                try{
                    aux2 = Double.valueOf(__valores[(int)i]).doubleValue();
                    if(aux.doubleValue() > aux2){
                        aux.setDouble(aux2);
                    }
                    soma = soma + aux2;
                    __n.setLong(__n.longValue() + 1);
                }catch(NumberFormatException e){
                    return(3);
                }
            }
        }
        __media.setDouble(soma/ __n.doubleValue());
        __minimo.setDouble(aux.doubleValue());
        return(0);
    }
}

//*****
// Acha o valor máximo da variável, além de calcular a média desta variável.
// Parâmetros:
//    __valores: valores da variável.
//    __n: retorna o número de casos da variável.
//    __media: retorna a média da variável.
//    __mínimo: retorna o valor máximo da variável.
//*****
public int AcheMaiorValor(String[] __valores, Objeto __n, Objeto __media, Objeto __maximo) {
    long i,tamanho;
    double aux2,soma;
    int ret;
    tamanho = __valores.length;
    Objeto aux = new Objeto(0);
    Objeto j = new Objeto(0);

```

```

ret = PrimeiroValor(__valores,aux,j);
if(ret !=0){
    return(ret);
}else{
    // aux contém o primeiro (não nulo) valor do array
    __n.setLong(1);
    soma = aux.doubleValue();
    for(i=j.intValue()+1; i<=tamanho-1;i++){
        if(!__valores[(int)i].equals("")){
            try{
                aux2 = Double.valueOf(__valores[(int)i]).doubleValue();
                if(aux.doubleValue() < aux2){
                    aux.setDouble(aux2);
                }
                soma = soma + aux2;
                __n.setLong(__n.longValue()+1);
            }catch(NumberFormatException e){
                return(3);
            }
        }
    }
    __media.setDouble(soma/__n.doubleValue());
    __maximo.setDouble(aux.doubleValue());
    return(0);
}
}

//*****
// Calcula o primeiro quartil da variável
// Parâmetros:
//     __valores: valores da variável.
//     __Q25: retorna o primeiro quartil da variável.
//*****
public int CalculeQ25(double[] __valores, Objeto __Q25){
    long n;
    int ret;
    Objeto _valores = new Objeto();
    _valores.__double = __valores;
    n = _valores.__double.length;
    if(n==0){
        return(8);
    }else{
        ret = OrdeneValores(_valores);
        if(ret != 0){
            return(ret);
        }else{
            __Q25.setDouble(_valores.__double[Math.round((n+1)/4)-1]);
            return(0);
        }
    }
}

//*****
// Calcula a mediana da variável
// Parâmetros:
//     __valores: valores da variável.

```

```
//      __mediana: retorna a mediana da variável.
//*****

public int CalculeMediana(double[] __valores, Objeto __mediana) {
    long n;
    int ret;
    Objeto _valores = new Objeto();
    _valores.__double = __valores;
    n = _valores.__double.length;
    if(n==0){
        return(8);
    }else{
        ret = OrdeneValores(_valores);
        if(ret != 0){
            return(ret);
        }else{
            if(n%2 == 0){ // É Par
                __mediana.setDouble((_valores.__double[(int)(n/2)-1]+__valores[(int)(n/2)])/2);
            }else{
                __mediana.setDouble(_valores.__double[(int)((n+1)/2)-1]);
            }
            return(0);
        }
    }
}

//*****
// Calcula o terceiro quartil da variável
// Parâmetros:
//      __valores: valores da variável.
//      __Q75: retorna o terceiro quartil da variável.
//*****

public int CalculeQ75(double[] __valores, Objeto __Q75) {
    long n;
    int ret;
    Objeto _valores = new Objeto();
    _valores.__double = __valores;
    n = _valores.__double.length;
    if(n==0){
        return(8);
    }else{
        ret = OrdeneValores(_valores);
        if(ret != 0){
            return(ret);
        }else{
            __Q75.setDouble(_valores.__double[Math.round(3*(n+1)/4)-1]);
            return(0);
        }
    }
}
}
```

Métodos exclusivos do Módulo de ACP

```
//*****
```

```

// Gera a matriz de correlação;
// Parâmetros:
//      __variaveis: é a matriz de dados das variáveis da base escolhida.
//*****
public double[][] makeCorrelacao(ObjetoRefVariaveisQuantitativas[] __variaveis) {
    int _nroVariaveis = __variaveis.length;
    int _nroCasos = __variaveis[0].variaveisQuantitativas.getValores().length;
    double[][] _matrizCorrelacao = new double[_nroVariaveis][_nroVariaveis];
    Double[][] _valores = new Double[_nroVariaveis][];
    for (int i = 0; i < _nroVariaveis; i++) {
        Double[] _val = new Double[_nroCasos];
        String[] _str = __variaveis[i].variaveisQuantitativas.getValores();
        for (int j = 0; j < _nroCasos; j++) {
            _val[j] = Double.valueOf(_str[j]);
        }
        _valores[i] = _val;
    }
    for (int i = 0; i < _nroVariaveis - 1; i++) {
        for (int j = 1; j < _nroVariaveis; j++) {
            long _soma = 0;
            long _soma2 = 0;
            long _soma3 = 0;
            long _soma4 = 0;
            long _soma5 = 0;
            for (int k = 0; k < _nroCasos; k++) {
                _soma += _valores[i][k].intValue();
                _soma2 += (_valores[i][k].intValue() * _valores[i][k].intValue());
                _soma3 += _valores[j][k].intValue();
                _soma4 += (_valores[j][k].intValue() * _valores[j][k].intValue());
                _soma5 += _valores[i][k].intValue() * _valores[j][k].intValue();
            }
            double _res1 = (_soma2 - (_soma * _soma) / _nroCasos);
            double _res2 = (_soma4 - (_soma3 * _soma3) / _nroCasos);
            double _res3 = _soma5 - (_soma * _soma3) / _nroCasos;
            double _resFinal = _res3 / (Math.sqrt(_res1 * _res2));
            _matrizCorrelacao[i][j] = _resFinal;
            _matrizCorrelacao[j][i] = _matrizCorrelacao[i][j];
        }
        _matrizCorrelacao[i][i] = 1.0;
    }
    _matrizCorrelacao[_nroVariaveis - 1][_nroVariaveis - 1] = 1.0;
    matrizCorrelacao = _matrizCorrelacao;
    return _matrizCorrelacao;
}

//*****
// Retorna a matriz de correlação, gerada no método makeCorrelacao();
// Parâmetros:
//*****
public double[][] getMatrizCorrelacao() {
    return matrizCorrelacao;
}

//*****

```

```

// Gera a tabela de Autovalores, AutoVetores, percentual total de Variância e o percentual
//      acumulado
// Parâmetros:
//      _matrizCorrelacaoEntrada: matriz de correlação das variáveis.
//      _variaveisCount: número de variáveis.
//*****
public void makeAnaliseEigen(double[][] _matrizCorrelacaoEntrada, long _variaveisCount) {
    double[][] _matrizCorrelacao = new double[(int)_variaveisCount][(int)_variaveisCount];
    for(int i = 0; i < (int)_variaveisCount; i++) {
        for(int j = 0; j < (int)_variaveisCount; j++) {
            _matrizCorrelacao[i][j] = _matrizCorrelacaoEntrada[i][j];
        }
    }
    double[][] _matrizAux = new double[(int)_variaveisCount][(int)_variaveisCount];
    double _d9 = 0;
    double[] _autoValores = new double[(int)_variaveisCount];
    for (int i = 0; i < _variaveisCount; i++) {
        for (int j = 0; j < _variaveisCount; j++) {
            _matrizAux[i][j] = 0.0;
        }
        _matrizAux[i][i] = 1.0;
    }
    double _c;
    double _aux = 0.00000001;
    double _soma = 0.0;
    for (int i = 1; i < _variaveisCount; i++) {
        for (int j = 0; j < i; j++) {
            _soma += 2*(_matrizCorrelacao[i][j]*_matrizCorrelacao[i][j]);
        }
    }
    double _t = Math.sqrt(_soma);
    double _verif = (_aux/_variaveisCount) * _t;
    boolean _passou = true;
    while (_passou) {
        double _d = 1;
        _t = _t/_variaveisCount;
        for (int k = 1; k < _variaveisCount; k++) {
            for (int l = 0; l < k; l++) {
                if (Math.abs(_matrizCorrelacao[l][k]) >= _t) {
                    _d = 1.0;
                    double _v = _matrizCorrelacao[l][l];
                    double _z = _matrizCorrelacao[l][k];
                    double _e = _matrizCorrelacao[k][k];
                    double _f = 0.5*(_v - _e);
                    double _g;
                    if (_f == 0)
                        _g = -1;
                    else
                        _g = -(getSign(_f));
                    _g = _g * _z / (Math.sqrt(_z*_z + _f*_f));
                    if (_g < -1)
                        _g = -0.999999;
                    double _h = _g / (Math.sqrt(2*(1 + Math.sqrt(1-_g*_g))));
                    double _k = Math.sqrt(1-_h*_h);
                    for (int m = 0; m < _variaveisCount; m++) {
                        if ((m == l) || (m == k)) {
                            _c = _matrizAux[m][l];
                            _f = _matrizAux[m][k];
                        }
                    }
                }
            }
        }
    }
}

```

```

        _matrizAux[m][k] = _c * _h + _f * _k;
        _matrizAux[m][l] = _c * _k - _f * _h;
    } else {
        _c = _matrizCorrelacao[m][l];
        _f = _matrizCorrelacao[m][k];
        _matrizCorrelacao[k][m] = _c * _h + _f * _k;
        _matrizCorrelacao[m][k] = _matrizCorrelacao[k][m];
        _matrizCorrelacao[l][m] = _c * _k - _f * _h;
        _matrizCorrelacao[m][l] = _matrizCorrelacao[l][m];
    }
}
_matrizCorrelacao[l][l] = _v*_k*_k + _e*_h*_h - 2*_z*_h*_k;
_matrizCorrelacao[k][k] = _v*_h*_h + _e*_k*_k + 2*_z*_h*_k;
_matrizCorrelacao[l][k] = (_v - _e) * _h*_k + _z * (_k*_k - _h*_h);
_matrizCorrelacao[k][l] = _matrizCorrelacao[l][k];
}
}
if (_t > 0) {
    _d = 0;
} else if (_d == 0) {
    _d = 0;
} else {
    _passou = false;
}
}
double[] _temp = new double[(int)_variaveisCount];
for (int i = 0; i < _variaveisCount; i++) {
    _temp[i] = i;
}
int _count = 0;
_passou = true;
while (_passou) {
    boolean _ok = false;
    _count++;
    for (int i = 0; i < (int)_variaveisCount - _count; i++) {
        if (_matrizCorrelacao[i][i] < _matrizCorrelacao[i+1][i+1]) {
            _ok = true;
            double _v2 = _matrizCorrelacao[i][i];
            _matrizCorrelacao[i][i] = _matrizCorrelacao[i+1][i+1];
            _matrizCorrelacao[i+1][i+1] = _v2;
            _v2 = _temp[i];
            _temp[i] = _temp[i+1];
            _temp[i+1] = _v2;
        }
    }
    if (!_ok)
        _passou = false;
}
double _q = 0;
double _d7 = 0;
double[] _percTotalVariancia = new double[(int)_variaveisCount];
double[] _percAcumVarianciaFator = new double[(int)_variaveisCount];
double[] _s3 = new double[(int)_variaveisCount];
double[][] _c1 = new double[(int)_variaveisCount][(int)_variaveisCount];
for (int i = 0; i < _variaveisCount; i++) {
    _autoValores[i] = _matrizCorrelacao[i][i];
    if (_autoValores[i] > 0) {
        _s3[i] = Math.sqrt(_autoValores[i]);
        _d7 += (_autoValores[i]);
    }
}

```

```

    } else {
        _autoValores[i] = 0.00000001;
    }
}
for (int i = 0; i < _variaveisCount; i++) {
    _percTotalVariancia[i] = 100 * _autoValores[i] / _d7;
    _d9 += (_autoValores[i])/_d7;
    _percAcumVarianciaFator[i] = 100.0 * _d9;
    int _k = (int)_temp[i];
    double _count2 = 0;
    for (int j = 0; j < _variaveisCount; j++)
        _count2 += _matrizAux[j][_k] * _matrizAux[j][_k];
    for (int j = 0; j < _variaveisCount; j++)
        _matrizCorrelacao[j][i] = _matrizAux[j][_k] * Math.sqrt(1.0/_count2);
}
for (int i = 0; i < _variaveisCount; i++) {
    _matrizCorrelacao[i][1] *= -1;
    _matrizCorrelacao[i][2] *= -1;
}

for (int i = 0; i < _variaveisCount; i++) {
    for (int j = 0; j < _variaveisCount; j++) {
        _c1[i][j] = 0.0;
        _matrizAux[i][j] = 0.0;
    }
    _c1[i][i] = _s3[i];
}
for (int i = 0; i < _variaveisCount; i++)
    for (int j = 0; j < _variaveisCount; j++)
        for (int k = 0; k < _variaveisCount; k++)
            _matrizAux[i][j] += _matrizCorrelacao[i][k] * _c1[k][j];
autoValores = _autoValores;
percTotalVariancia = _percTotalVariancia;
percAcumVarianciaFator = _percAcumVarianciaFator;
autoVetores = _matrizCorrelacao;
compLoadings = _matrizAux;
makeAutoVetorPercVariancia((int)_variaveisCount);
}

```

```

//*****
// Retorna a tabela de Autovalores, calculada no método makeAnaliseEigen();
// Parâmetros:
//*****
public double[] getAutoValores() {
    return autoValores;
}

```



```

//*****
// Retorna a tabela de Autovetores, calculada no método makeAnaliseEigen();
// Parâmetros:
//*****

    public double[][] getAutoVetores() {
        return autoVetores;
    }

//*****
// Retorna a tabela de percentuais totais de Variância, calculada no método
//      makeAnaliseEigen();
// Parâmetros:
//*****

    public double[] getPercTotalVariancia () {
        return percTotalVariancia;
    }

//*****
// Retorna a tabela de percentuais acumulados de Variância, calculada no método
//      makeAnaliseEigen();
// Parâmetros:
//*****

    public double[] getPercAcumVarianciaFator() {
        return percAcumVarianciaFator;
    }

```